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 #:** 0018

**Corresponding Measures:**

**De.2. Measure Title:** Controlling High Blood Pressure

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

**De.3. Brief Description of Measure:** The percentage of adults 18-85 years of age who had a diagnosis of hypertension (HTN) and whose blood pressure was adequately controlled (<140/90 mm Hg) during the measurement year.

**1b.1. Developer Rationale:** One out of every three Americans have hypertension, or high blood pressure. Even with the availability of effective treatment options, only about half (54%) of these people have their high blood pressure under control (Merai et al, 2016). Improvements in quality or better control of blood pressure as related to this measure would help significantly reduce the probability of serious and costly complications, including coronary artery disease, congestive heart failure, stroke, ruptured aortic aneurysm, renal disease and retinopathy.

Merai R, Siegel C, Rakotz M, Basch P, Wright J, Wong B; DHSc., Thorpe P. CDC Grand Rounds: A Public Health Approach to Detect and Control Hypertension. MMWR Morb Mortal Wkly Rep. 2016 Nov 18;65(45):1261-1264.

**S.4. Numerator Statement:** Patients whose most recent blood pressure level was <140/90 mm Hg during the measurement year.

**S.6. Denominator Statement:** Patients 18-85 years of age who had at least two visits on different dates of service with a diagnosis of hypertension during the measurement year or the year prior to the measurement year.

**S.8. Denominator Exclusions:** This measure excludes adults in hospice. It also excludes adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings.

Additionally, this measure excludes patients with evidence of end-stage renal disease, dialysis, nephrectomy, or kidney transplant on or prior to the December 31 of the measurement year. It also excludes female patients with a diagnosis of pregnancy during the measurement year, and patients who had a nonacute inpatient admission during the measurement year.

**De.1. Measure Type:** Outcome: Intermediate Clinical Outcome

**S.17. Data Source:** Claims, Electronic Health Data, Electronic Health Records, Paper Medical Records

**S.20. Level of Analysis:** Health Plan

## Preliminary Analysis: Maintenance of Endorsement

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

### Criteria 1: Importance to Measure and Report

#### 1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

**Evidence Summary**

- The developer provides a diagram outlining the steps between the process and the intermediate outcome (adequate control of blood pressure), and how the intermediate outcome in turn influences the longer-term outcomes (reduction in cardiovascular events).
- The evidence base for this measure includes two graded clinical practice guidelines, one from the American College of Cardiology (ACC)/American Heart Association (AHA) and one from the American College of Physicians (ACP) and the American Academy of Family Physicians (AAFP). The guidelines differ in age of target population and recommend different blood pressure goals.
    - ACC/AHA Recommendation 1: "For adults with confirmed hypertension and known CVD or 10-year ASCVD event risk of 10% or higher, a BP of less than 130/80 mm HG is recommended" Class I; Level B-R (systolic), Level C-EO (diastolic)
    - ACC/AHA Recommendation 2: "For adults with confirmed hypertension, without additional markers of increased CVD risk, a BP target of less than 130/80 mm HG may be reasonable" Class IIb; Level B-NR (systolic), Level C-EO (diastolic)
    - ACP/AAFP Recommendation 1: "ACP and AAFP recommend that clinicians initiate treatment in adults aged 60 years or older with systolic blood pressure persistently at or above 150 mm Hg to achieve a target systolic blood pressure of less than 150 mm Hg to reduce the risk for mortality, stroke, and cardiac events. (Grade: strong Recommendation, high-quality evidence). ACP and AAFP recommend that clinicians select the treatment goals for adults aged 60 years

or older based on a periodic discussion of the benefits and harms of specific blood pressure targets with the patient."

- ACP/AAFP Recommendation 2: "ACP and AAFP recommend that clinicians consider initiating or intensifying pharmacologic treatment in adults aged 60 years or older with a history of stroke or transient ischemic attack to achieve a target systolic blood pressure of less than 140 mm Hg to reduce the risk for recurrent stroke. (Grade: weak recommendation, moderate-quality evidence). ACP and AAFP recommend that clinicians select the treatment goals for adults aged 60 years or older based on a periodic discussion of the benefits and harms of specific blood pressure targets with the patient."
- ACP/AAFP Recommendation 3: "ACP and AAFP recommend that clinicians consider initiating or intensifying pharmacologic treatment in some adults aged 60 years or older at high cardiovascular risk, based on individualized assessment, to achieve a target systolic blood pressure of less than 140 mm Hg to reduce the risk for stroke or cardiac events. (Grade: weak recommendation, low-quality evidence). ACP and AAFP recommend that clinicians select the treatment goals for adults aged 60 years or older based on a periodic discussion of the benefits and harms of specific blood pressure targets with the patient."

**Changes to evidence from last review**

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

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

**Updates:** All evidence provided is updated since the last review by the standing committee.

***Questions for the Committee:***

- The evidence provided by the developer is updated since the previous NQF review. Does the Committee agree there is a need for discussion and vote on Evidence?

**Guidance from the Evidence Algorithm**

Intermediate outcome measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: moderate; Consistency: high (Box 5) → Moderate (Box 5b) → Moderate

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

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

**Maintenance measures – increased emphasis on gap and variation**

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

- Developer presents the most recent years of results (from HEDIS) for this measure stratified by product line (commercial, Medicaid, Medicare). Results are at the health plan level.

Commercial Health Plans

| Year | # of plans | Mean | St Dev | Min | 10th | 25th | 50th | 75th | 90th | MAX | Inter-quartile Range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2016 | 352 | 59% | 12% | 22% | 44% | 50% | 57% | 68% | 75% | 90% | 18% |
| 2017 | 350 | 58% | 12% | 26% | 44% | 51% | 58% | 66% | 75% | 92% | 15% |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2018 | 403 | 55% | 21% | 0% | 9% | 52% | 60% | 67% | 74% | 85% | 15% |

Medicaid Health Plans

| Year | # of plans | Mean | St Dev | Min | 10th | 25th | 50th | 75th | 90th | MAX | Inter-quartile Range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2016 | 259 | 56% | 12% | 25% | 40% | 48% | 57% | 65% | 72% | 90% | 17% |
| 2017 | 264 | 57% | 13% | 2% | 42% | 49% | 59% | 66% | 71% | 85% | 17% |
| 2018 | 248 | 59% | 13% | 0% | 46% | 53% | 61% | 67% | 72% | 85% | 14% |

Medicare Health Plans

| Year | # of plans | Mean | St Dev | Min | 10th | 25th | 50th | 75th | 90th | MAX | Inter-quartile Range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2016 | 466 | 70% | 13% | 24% | 51% | 61% | 71% | 81% | 84% | 97% | 19% |
| 2017 | 460 | 71% | 14% | 8% | 54% | 63% | 74% | 80% | 86% | 96% | 18% |
| 2018 | 483 | 69% | 11% | 0% | 57% | 64% | 71% | 76% | 81% | 100% | 12% |

**Disparities**

- The developer states they do not currently collect performance data stratified by race, ethnicity, or language.
- The developer provides summary data from a report produced by the CMS Office of Minority Health in collaboration with RAND Corporation. The report looks at disparities in the Medicare Advantage population. The report infers race and ethnicity and reports the following results:
    - White and Asian or Pacific Islander: 69% control rate for hypertension
    - Black: 59% control rate for hypertension
    - Hispanic: 67% control rate for hypertension
- The developer summarizes literature demonstrating variation in the prevalence of hypertension by race and that there are disparities in awareness, treatment, and control of hypertension.

***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**

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability: Specifications and Testing**

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

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

## Reliability

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

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

## Validity

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

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

**Composite measures only:**

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

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

**Evaluators:** NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

**Methods Panel Evaluation Summary**:

This measure was reviewed by the Scientific Methods Panel and discussed during the October 2019 In-Person Meeting. The Subgroup passed the measure on reliability. The Subgroup was unable to reach consensus regarding validity. The full Panel discussed validity and then the Subgroup revoted, passing the measure on validity. A summary of the measure and the Panel discussion is provided below.

Reliability: 4-H; 1-M; 0-L; 2-I

- Reliability of the health plan measure score was tested using a beta-binomial approach (i.e., signal to noise); Overall reliability ranged 0.982-0.999 across the three types of plans
- Reviewers expressed concerns with clarity and consistency of specifications (i.e., different age ranges used throughout the specifications, lack of clarity around target blood pressure, inconsistencies in the denominator and numerator details)
- After the initial review, the developer responded to the reviewers' concerns and updated materials for the measure to clarify the specifications.

Validity: 0-H; 4-M; 2-L; 0-I

- Validity of the health plan measure was demonstrated through construct validity using the entire HEDIS data sample.
    - Construct validity of the Controlling Blood Pressure measure was conducted by assessing the correlation with another measure: Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg): The percentage of adults 18-75 years of age with diabetes (type 1 and type

2) whose most recent blood pressure level taken during the measurement year is <140/90 mm Hg.
    - Pearson correlation across the three types of health plans ranged 0.75 to 0.93; Medicare with the lowest and commercial plans with the highest correlation score.
- Concerns with validity included lack of analysis around multiple data sources, analysis of exclusions, lack of risk adjustment, and the measure used to correlate for construct validity. The Subgroup was initial unable to reach consensus on validity.
- After the initial review, the developer responded to the reviewers' concerns and provided updated materials. On re-vote, the measure passed validity with a moderate rating.
    - The developer hypothesized that health plans that perform well managing one chronic condition (hypertension) should perform well managing other chronic conditions. They repeated the construct validity analysis using two a1C control measures: NQF #0575 *Comprehensive Diabetes Care: HbA1c Control (< 8%)* and NQF #0059 *Comprehensive Diabetes Care: HbA1c Poor Control (> 9%)*.
        - Pearson correlation with #0575 across the three types of health plans ranged from 0.51 to 0.81; Medicare with the lowest and commercial with the highest correlation score.
        - Pearson correlation with #0059 across the three types of health plans ranged from -0.58 to -0.82; Medicare with the lowest correlation score and commercial and Medicaid very similar.
    - The Scientific Methods Panel discussed the lack of risk adjustment and whether that allowed for adequate comparison of health plans. Intermediate outcome measures are not required to include risk adjustment under NQF endorsement criteria.

***Questions for the Committee regarding reliability:***

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

***Questions for the Committee regarding validity:***

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

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

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

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

2a1.

- Fairly high reliability overall except for the inconsistent age range citations.
- The measure was rated and passed as high by the sci methods panel
- Most recent blood pressure setting is not defined, I assume that data gathering is only from discrete BP data fields, not free text entries.
- High per methods panel

- Reviewed by Scientific Methods Panel. Passed reliability.
- The measure introduces new exclusions for frailty, advanced illness, and dementia. It's unclear how well this will perform. The inclusion of remote BP measurements is a positive change. Measure score testing was done using signal to noise ratio showed strong reliability.

2a2.

- My only concerns about the reliability are the unclear age range and the selection method for "most recent blood pressure", which seems murky.
- No
- Office blood pressures usually higher than real world, but no way to fix this
- High per methods panel
- No.
- No

2b1.

- Some of the exclusions do not appear well justified (e.g., pregnancy) and/or difficult to accurately identify (frailty/debilitating illness)..
- no concerns. I endorse the conclusion of the SMP
- not sure why patients with nonacute inpatient admissions excluded
- Moderate per methods panel. Correlation between chronic condition control measures lower for MA plans and highest for commercial plans (Medicaid in between)
- Risk adjustment would be beneficial. Additionally, more information regarding the target BP in relation to age may impact validity.
- Construct validity testing was done with two measures for HbA1c control. Correlations were near 0.8 overall, though in the 0.5-0.6 range for Medicare health plans

2b4-7.

- No
- I don't see any significant threats to validity
- possibility for systemic bias against patients with paper records
- None
- No concerns.
- I didn't find much on missing data

2b2-3

- Risk adjustment is not included, and its advisability should be discussed in the in-person meeting.
- The measure is not risk adjusted
- It is inappropriate to exclude BPs taken or reported by patient. Patients wiht well controlled BP who do not have an in person visit in the measurement period will be inappropriately excluded.
- No risk adjustment for race even though lower SDS is associated with lower plan performance
- A deeper review may be warranted by the committee. This was an area the Scientific Methods discussed.
- exclusions with different age ranges could be a threat. The measure is not risk adjusted.

## Criterion 3. Feasibility

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

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

- Data elements are routinely generated and used during care delivery (e.g. blood pressure). Data elements are generally available in electronic form, such as EHR and claims data. If providers participating in the health plan are using paper records, the health plan may need to abstract data from paper records for those providers.

***Questions for the Committee:***

- Do you have any concerns with the feasibility of this measure?

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

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

- Feasbility appears solid.
- Providers who do not have an EHR will need to manually abstract BP data
- possibility for systemic bias against patients with paper records
- No issues
- No concerns
- Data elements can be captured through routine care

## Criterion 4: Usability and Use

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

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

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

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

**Current uses of the measure**

**Publicly reported?** ☒ **Yes** ☐ **No**

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

**OR**

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

**Accountability program details**

The developer reports the measure is used in the following accountability programs:

Public Reporting

- Health Plan Ratings

- Health Plan Report Card

Payment Program

- CMS Medicare Star Rating Program
- CMS Medicaid Adult Core Set
- CMS Quality Payment Program
- California's Value Based Pay for Performance Program

Regulatory and Accreditation Programs

- NCQA Accreditation

Quality Improvement (external benchmarking to organizations)

- Quality Compass
- Annual State of Health Care Quality

**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**

Health plans calculate the measure using either their own data or data they have collected. They have immediate access to their results. The developer publicly reports results annually and provides benchmarks to place results into context.

The developer provides technical assistance on measures through their Policy Clarification Support System. In addition to input obtained through the technical assistance process, the developer states they obtain input through multi-stakeholder advisory panels and public commenting. During the last major update for the measure, feedback obtained through these methods informed several changes to the measures:

- Reworking the denominator approach to reduce burden,
- Adding an administrative approach for the numerator,
- Including readings from remote monitoring devices, and
- Updating the numerator threshold to focus on a target of <140/90 mm Hg

***Questions for the Committee**:*

- How have the performance results been 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 states performance has been generally improving over the past several years with each plan type demonstrating performance improvement each year of around 1%. The measure changes made in 2018 (see feedback section above) broke this trend.

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

**Unexpected findings (positive or negative) during implementation**

The developer states there have been no unexpected findings.

**Potential harms**

No information included on potential harms.

**Additional Feedback:**

***Questions for the Committee**:*

- Are you aware of any potential harms not identified here?
- Do the benefits of the measure outweigh any potential unintended consequences?

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

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

4a1.

- My only concern about use is the apparent lack of improvement in the past three years of the measure's implementation. The performance gap is still quite large. Is this measure working? If not, why not?
- The measure is publicly reported and currently used in accountability programs.
- Yes
- No issues
- Has major uptake in reporting programs. Steward has received feedback.
- It is currently publicly reported.

4b1.

- It is unclear how the implementation of this measure could be improved, to improve the BP control outome, and thus reduce stroke and heart attacks.
- Performance is improving slowly. Benefits outweigh unintended consequences
- May force otherwise unneeded office visits to document BP, potential for economic harm to patient and overall system waste.
- None
- No issues
- demonstrated usability

## Criterion 5: Related and Competing Measures

**Related or competing measures**

0061 Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)
2602 Controlling High Blood Pressure for People with Serious Mental Illness

2606 Diabetes Care for People with Serious Mental Illness: Blood Pressure Control (<140/90 mm Hg)
0729 Optimal Diabetes Care (Minnesota Community Measurement)
0076 Optimal Vascular Care (Minnesota Community Measurement)

**Harmonization**

The measures are harmonized to the extent possible, with the possible exception of 2602. In 2602, for a subset of patients (ages 60-85 without a diagnosis of diabetes), the BP target is <150/90 mm Hg. Some of those patients may also qualify for the denominator of 0018, yielding conflicting BP targets.

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

- 0061 Diabetes BP control, 2602 BP Control with Serious Mental Illness, 2606 Diabetes Care for People with Serious Mental Illness--BP control
- In 2602, for a subset of patients (ages 60-85 without a diagnosis of diabetes), the BP target is <150/90 mm Hg. Some of those patients may also qualify for the denominator of 0018, yielding conflicting BP targets.
- would be nice to have same BP goal, but clinical data is conflicting.
- Generally yes with targets <140/90
- The steward made note of the conflicting systolic BP goals in geriatric patients.
- Some of the related measures use different BP cutoffs.

## Public and Member Comments

**Comments and Member Support/Non-Support Submitted as of: Month/Day/Year**

- **Of the XXX NQF members who have submitted a support/non-support choice:**
  - XX support the measure
  - YY do not support the measure

### Combined Methods Panel Scientific Acceptability Evaluation

**Type of measure:**

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

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

**Data Source:**

☒ **Claims** ☒ **Electronic Health Data** ☒ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☒ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☐ **Enrollment Data** ☒ **Other**

**Panel Member #1:** marked under S.23 but not described

**Level of Analysis:**

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

☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

**Measure is:**

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

**RELIABILITY: SPECIFICATIONS**

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

   **Submission document:** "MIF_0018" document, items S.1-S.22

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

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

   **Panel Member #1:** S.6. Numerator Details:

   "The patient is not compliant if the blood pressure is =140/90 mm Hg".

   "The patient is not compliant if the BP reading is =140/90 mm Hg or is missing…"

   I assume there is a typo – i.e., ≥?

   There are exclusion criteria described under the numerator details that are not included in the denominator exclusions. For example:

   "Do not include BP readings:

   - Taken during an acute inpatient stay or an ED visit.

   - Taken on the same day as a diagnostic test or diagnostic or therapeutic procedure that requires a change in diet or change in medication on or one day before the day of the test or procedure, with the exception of fasting blood tests.

   - Reported by or taken by the patient."

   It is preferred that all exclusions are detailed within the S.10. Denominator Exclusions section, e.g., Two telehealth visits, Type of visits (Diagnostic tests), self-report.

S.11. Denominator Exclusion Details

**Panel Member #1:** In addition to the comments above, some age ranges are described over the age of 75. Consistency with the max age allowed throughout the submission is preferred.

**Panel Member #2:** The brief description of measure (De.3.) gives the denominator as adults 18-85 and the denominator statement (S.7.) gives the denominator as 18-75. In addition, the denominator exclusion details (S.11.) mentions an exclusion for adults 66-80. Also there are several out-of-date dates (Ad.2-Ad.5) which raises the concern of general accuracy of the document.

**Panel Member #5:** a)The Brief Description of Measure (C01.1) says 18-85 year of age. Other documentation says 18-75.

b)S.10. The following is listed as an exclusion "patients who had a nonacute inpatient admission during the measurement year" but no rationale is given.

**Panel Member #6:** *There are many potential sources of data for the "whose most recent blood pressure level," and since BP readings are not exact, it seems possible that different "most recent" measures, depending on how one looks for them, might be different. It also seems possible that there could be*

*variability in whether the patient meets the 140/90 threshold (see MIF top of p. 8). I could also imagine similar problems with the denominator. There have been a number of changes in the specifications since the last cycle, so this needs to be addressed with the new specifications.*

**Panel Member #7:** The denominator is limited to patients 18-75 years of age, but the denominator exclusions pertaining to administrative claims analysis refer to scenarios with adults 66-80 years of age and with adults 81 years of age and older. I think that any references to exclusion of patients 76 years of age and older is confusing, given the denominator specification.

**RELIABILITY: TESTING**

**Submission document:** "MIF_0018" 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**

   **Panel Member #1:** In section 1.6, it may be useful to explain why and how a sample of 411 medical records per plan was selected for testing. If all plans assessed exactly 411 medical records, why was the median reported and not simply the number of records? It would be useful to describe how many cases were assessed including claims and medical records per product line, e.g., count, mean, median, min, max.

   **Panel Member #3:** Used beta-binomial test to assess SNR.

   **Panel Member #6:** *Although either Measure score OR Data element testing is required, in this case I am concerned that Measure score testing alone does not address the measure specification concerns identified above. I would like to see some testing in which the numerator and denominator are independently reassessed by a different person/method.*

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**

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

   **Submission document:** Testing attachment, section 2a2.2

   **Panel Member #1:** The methods used for score reliability testing cannot be fully assessed. Beta-binomial model (Adams, 2009), but no details were provided on the actual methods and formulas used.

   More details on the statistical method and specific formulas used to calculate the proportion of variability in measured performance that can be explained by real differences in performance would be helpful to better understand what was done exactly.

   **Panel Member #2:** The developer uses a common approach the Beta-binomial model (Adams 2009) to estimate signal-to-noise and reports the reliability statistic distribution (although not stratified by volume which is likely because all plans report the same number of cases = 411).

   **Panel Member #3:** Used beta-binomial test to assess SNR.

   **Panel Member #4:** *Reliability testing was at the appropriate level of analysis (a few hundred patients within a few hundred health plans, stratified by product line (commercial, Medicare, Medicaid) using beta-binomial calculation, a method described by RAND, among others, for dichotomous outcomes.*

   **Panel Member #6:** *For Measure score level testing, the Beta-binomial is appropriate.*

**Panel Member #7:** The testing is appropriate. The steward fit a beta-binomial regression model to the plan-level measure values, thus permitting estimation of beta distribution parameters. From this, reliability estimates can be readily derived, according to well-known statistical methodology.

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

**Submission document:** Testing attachment, section 2a2.3

**Panel Member #1:** The results are clearly demonstrated in table 2.

Some additional information would be helpful, for example:

- What was the minimum number of patient records needed to be assessed by a health plan to reach the recommended reliability threshold of 0.7? These reliability results may be strongly impacted by the number of cases assessed within each heal plan, which seems to be a constant number of 411, but I'm not sure. Therefore, knowing the minimum required number of cases needed to achieve acceptable levels of reliability is an important information to have.
- The counts per product line would be a nice addition to this table for clarity.

**Panel Member #2:** The developer reports that the reliability for most of the health plans across product lines demonstrate high reliability, with perhaps a small number of Medicaid health plans demonstrating low to moderate reliability.

**Panel Member #4:** *Reliability was quantified as >0.9) within each "product line."*

**Panel Member #5:** Overall and plan-level reliabilities were very high, based on a random sample of ~414 patients per plan.

**Panel Member #6:** *The results strongly demonstrate reliability.*

**Panel Member #7:** Reliability appears to be excellent. Mean reliability estimates for each of commercial, Medicaid, and Medicare health plans exceed 0.98. There is potential for unreliable estimates in a very small share of Medicaid health plans.

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

**Submission document:** Testing attachment, section 2a2.2

☒ **Yes**

**Panel Member #1:** The developer reports that the reliability for most of the health plans across product lines demonstrate high reliability, with perhaps a small number of Medicaid health plans demonstrating low to moderate reliability.

☐ **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)

**Panel Member #6:** *But see above about the need for Data element testing*

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

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

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

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

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

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

**Panel Member #1:** The 'insufficient' rating was selected because of the lack of information on how the sample size per plan was selected, as it may strongly impact the reliability results.

**Panel Member #2:** For the level of reporting (i.e. health plan by product line with a standard sample size) the reliability must be rated as high, with perhaps the exception of some Medicaid plans). Reporting reliability results using an alternative methodology (e.g. ICC) must increase confidence in the application of Beta-binomial model since several of the health plans demonstrate perfect reliability (a metric of 1.0) which seems unlikely with a fixed denominator.

**Panel Member #3:** Score-level reliability was excellent.

**Panel Member #4:** *Despite very high statistical reliability, my concern, which may be unfounded, is that for an unadjusted intermediate clinical outcome, variation due to clinical and SES factors is falsely attributed to plan, resulting in an inflation of reliability. The measure developers thoughtfully offer than restricting reporting to within a single product line (Medicaid only, for example), offers some meaningful degree of SES adjustment. The value of adjustment for SES for this measure is beyond the scope of this review. The provided materials describe an exploration of the role of SES on plan rank, however, technical details were not to be found.*

**Panel Member #5:** The reliability data is compelling.

The specifications allow for administrative and/or EHR extracted data for numerator specifications. The application states that plan usually have a mix of methods in their data. No analyses were conducted to check how reliability might be affected by mode of data collection.

**Panel Member #6:** *My concern is based on the lack of Data element level testing.*

**Panel Member #7:** Greater than 90% of health plans in each strata (commercial, Medicaid, Medicare) have reliability exceeding 0.9. This can be expected, given that the measure is a simple proportion based on a random sample of 411 subjects. In other words, within-plan measurement error is limited by sample size.

**VALIDITY: ASSESSMENT OF THREATS TO VALIDITY**

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

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

**Panel Member #1:** Face validity established for setting the exclusion criteria seem strong and clinically logical.

The concern I had was the ability to correctly identify the excluded patients, using available data from all different health plans.

Developers tested the feasibility and impact of applying these exclusions to the measure, and found that they were feasible and excluded on average 3.8% of the sample. However, no sensitivity analyses were reported to assess the accuracy of the excluded sample as to the reasons for their exclusions. Could it be that a significantly larger sample had to be excluded, but were not excluded due to missing data points related to any of the large amount of exclusion criteria?

It is reported that: "NCQA compared several approaches for identifying the advanced illness and frailty populations, examining different age ranges and diagnosis positions and their impact on the denominator size. The results of those queries along with input from the expert work groups, measurement advisory panels and public comment led us to determine that the best approach for identifying the advanced illness and frailty population that should be excluded from the measure was to apply the following criteria:…"

Could the developers elaborate on the results mentioned above, and how they led them to determine the final criteria?

Could there also be a way to establish a range of valid exclusion rates using existing literature, as an additional method to assess the validity of the 3.8% excluded sample? In other words, given the exclusion criteria, what is the expected exclusion rate?

**Panel Member #2:** In general the stated exclusions are well-rationalized and the lengthy discussion of advanced illness and frailty seems well justified and the percent excluded (~4%) small. As always more data are better and reporting of the rate stratified by with and without the exclusion would be informative.

**Panel Member #3:** none

**Panel Member #4:** *I confess some ignorance here:*

*I would appreciate learning the rationale for*

**including a patient with one office visit followed by one telephone visit.*

**excluding patients with renal disease/nephrectomy manifesting by Dec 31 of the assessment year and the rationale,*

** excluding pregnant women as blood pressure management is clinically important for many of these patients (perhaps there is a different measure for such patients or determination of pregnancy is problematic.)*

**Panel Member #5:** In general, the method for determining exclusions was principled and systematic

S.10. The following is listed as an exclusion "patients who had a nonacute inpatient admission during the measurement year" but no rationale is given.

**Panel Member #6:** *None.*

**Panel Member #7:** My single concern is that some of the listed exclusions do not pertain to the measure, which is already limited to patients 18-75 years of age.

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

**Submission document:** Testing attachment, section 2b4.

**Panel Member #1:** No concerns. I suggest a minor edit by modifying the interpretation, to claim only the identification of statistically significant differences, not meaningful differences, as meaningfulness of these differences were not assessed as far as I can tell.

**Panel Member #2:** Given how long this measure has been in use the more relevant question might be the ability to identify meaningful differences in performance over time. Is there any evidence that health plans are able to improve the quality of care?

**Panel Member #3:** none

**Panel Member #4:** *Please note question above re: potential for lack of risk adjustment to inflate reliability estimate.*

**Panel Member #5:** None

**Panel Member #6:** *Although it would have been good to put the three box plots on pp. 13-14 in a single graph, the methods are appropriate and results positive.*

**Panel Member #7:** I have no concerns.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
**Submission document:** Testing attachment, section 2b5.

**Panel Member #1:** NA
**Panel Member #2:** Not applicable (although one might wonder whether the administrative data and/or medical chart review is relevant).
**Panel Member #3:** none
**Panel Member #4:** *Only one set of measure specifications.*
**Panel Member #5:** The specifications allow for administrative and/or EHR extracted data for numerator specifications. The application states that plan usually have a mix of methods in their data. No analyses were conducted to check how reliability might be affected by mode of data collection.
**Panel Member #6:** *As noted in Q #2, this may be a reliability concern.*
**Panel Member #7:** This portion of the form is not completed. I am confused by this, because it seems to that administrative claims and medical records are fundamentally different. My understanding is that the former draws upon CPT codes that document blood pressure ranges, rather than actual blood pressure measurements.

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

**Submission document:** Testing attachment, section 2b6.
**Panel Member #1:** There is a general description about audits that:
"verify primary data sources used to populate measures and ensure specifications are correctly implemented."
"If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used."

No specific results are provided about rates of missing data and number of health plans excluded due to missing data. I suggest this information be added. If the rate of excluded plans is not negligible, some testing on threats to validity due to these exclusions is warranted.

**Panel Member #2:** The HEDIS data are subject to systematic audit which is one of the advantages of the HEDIS measurement system.

**Panel Member #3:** none
**Panel Member #4:** *None.*
**Panel Member #6:** *No concerns.*
**Panel Member #7:** I have no concerns.

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

**Panel Member #3:** They do not present a strong rationale to justify lack of risk adjustment. A priori, it would seem that BP control would be more difficult to achieve in some patient populations (e.g. elderly patients, patients with multiple comorbidities) compared to others. There may also be a rationale for

maintaining higher BP in some clinical groups to maintain adequate end-organ perfusion (e.g. people with prior strokes).

**Panel Member #4:** *This is subject to debate.*

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

**Panel Member #1:** A rationale is provided for not risk adjusting, using a conceptual reasoning and an example.

The conceptual reasoning is the potential to mask poor performance and disparities in care. The example provided relates to socioeconomic status (SES). This example, also supported by testing results of no impact of adjusting for SES for this measure, provides sufficient justification for not adjusting for SES.

However, I do not think it provides sufficient justification for not adjusting for other patient characteristics, e.g., age, weight, etc. Although there could be justified reasons for not adjusting for patients' health and demographic characteristics other than SES, such justifications were not discussed or tested. Therefore, it is difficult to no risk-adjustment without doing so. There is a wide range between no adjustment and over adjustment, that should be further discussed and tested.

**Panel Member #2:** Given the amount of variation in performance across plan type (esp. Medicaid) and across health plans shown in Table 4 it is difficult to conclude that all of that variation is attributable to quality of care (and also the high estimates of reliability) rather than patient factors or other contextual factors (urban or rural).

**Panel Member #3:** This measure is not risk-adjusted. I would highly recommend that the measure developer add risk adjustment.

**Panel Member #5:** The decision to not risk adjust is sound**.**

**Panel Member #7:** The measure is not risk-adjusted. The materials indicate that risk-adjustment for low-income status, dual eligibility, and disability had no substantial impact on plan ranks. Ultimately, I have no concern about the absence of risk adjustment, as blood pressure is a known risk factor for mortality and morbidity.

**<u>For cost/resource use measures ONLY</u>:**

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

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

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

**VALIDITY: TESTING**

19. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**
20. **Method of establishing validity of the measure score:**
    - ☐ **Face validity**
    - ☒ **Empirical validity testing of the measure score**
    - ☐ **N/A (score-level testing not conducted)**
21. **Assess the method(s) for establishing validity**

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

    **Panel Member #1:** I found the validity testing provided somewhat challenging.

    If I understand correctly, plan level scores for the blood pressure measure were correlated with the same data element measure (blood pressure<140/90 mm Hg) for a sub-group of patients also included in the denominator. If this is correct, since the same data element measures were used for both groups, this is somewhat of a circular test of the same data element measure for basically two overlapping samples. Consequently, high correlations are expected, but these high correlations do not establish an empirical/predictive validity of this measure. Information on the extent or this overlap would help assess this potential bias. Developers could also test this same correlation with no overlap in samples, which would make a 'cleaner' and unbiased test.

    Construct validity at the score level could be to test a hypothesis that health plan scores would differ in clinically logical ways between subgroups of patients included in the denominator. For example, a health plan could be expected to have different scores when assessing patients with or without diabetes. If such hypothesis is supported in the expected direction, this strengthens the confident that the measure is capturing high blood pressure as expected.

    **Panel Member #2:** The developer examines the Pearson correlation among related measures (blood pressure control in two related conditions hypertension and diabetes). The stated implicit quality construct is "blood pressure control".

    **Panel Member #3:** Examined construct validity by examining the correlation of this measure with a separate measure evaluating BP control in patients with diabetes. Results show high correlation. This assessment is not very strong since the diabetes population is essentially a subset of the overall population. They are essentially assessing exactly the same measure in a subset of the population.

    **Panel Member #4:** *The provided materials evaluate construct validity by comparing this measure to a similar measure specific to patients with diabetes and find a Pearson correlation of about 0.75 or greater depending on product line (Medicare, Medicaid, Commercial.)*

    **Panel Member #5:** A Pearson correlation between this measure and a very similar measure for blood pressure control among patients with diabetes was done. It is unclear is the samples for these measures were independent. For example, if a patient had both HTN and DM, could they be included in both measures? If so, it would be surprising if the measures were not highly correlated.

    **Panel Member #6:** *Construct validity was assessed by examining the correlation between the proposed measure and another closely related measure: The percentage of adults 18-75 years of age with diabetes (type 1 and type 2) whose most recent blood pressure level taken during the measurement year is <140/90 mm Hg. These two measures would be expected to be highly correlated because (a) the numerator is essentially the same and (b) there is a large overlap in the denominator (patients with both hypertension and diabetes).*

    **Panel Member #7:** The materials indicate that correlations of this measure and the percentage of patients with diabetes who have controlled blood pressure (likewise, <140/90 mm Hg) were estimated. Correlations were very high in commercial and Medicaid health plans, but modestly lower than in Medicare health plans.

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

**Submission document: Testing attachment, section 2b2.3**

**Panel Member #1:** Due to the concerns raised above, the correlation presented do not necessarily provide evidence that the measure is an indicator of quality

**Panel Member #2:** The developer reports correlation among component measures in excess of 0.75)-0.93 across plan types at the measured entity level.

**Panel Member #3:** Correlation coefficient was very high.

**Panel Member #4:** The provided assessment is imperfect, of course, but fair.

**Panel Member #5:** The correlations between the measures was very high (>0.75). This provides evidence of concurrent validity., although it is unclear if the samples were independent.

**Panel Member #6:** *Given the expected high correlation (see Q #21), the results (rho = 0.93, 0.89, and 0.75) are not particularly impressive or indicative of high validity.*

**Panel Member #7:** The approach is reasonable, but it remains unclear whether the measure correlates with the percentage of patients without diabetes who have controlled blood pressure.

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

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

☒**Yes**

☒ **No**

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

**Panel Member #3:** I graded this as a "weak" yes for the reasons described above.

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

*NOTE that data element validation from the literature is acceptable.*

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

☐ **Yes**

☐ **No**

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

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

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

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

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

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

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

**Panel Member #1:** Validity testing may not be appropriate. Additional information on the potential circularity of the correlation test is needed.

Exclusion accuracy not assessed.

No risk-adjustment was not sufficiently justified.

**Panel Member #2:** A demonstration of an implicit quality construct is the lowest level of empirical validity testing. To demonstrate a moderate level, the developer must show an empirical association between the implicit quality construct and the material outcome (or better yet an explicit quality construct and the material outcome). For example, that health plans with worse performance on controlling high blood pressure among hypertension / diabetes patients have worse performance on coronary artery disease, congestive heart failure, stroke, ruptured aortic aneurysm, renal disease and retinopathy.

**Panel Member #3:** See above – due to limitations of assessment of construct validity.

**Panel Member #4:** *Reasonable given necessary limitations.*

**Panel Member #5:** Although generally strong, I am concerned that the samples used in the 2 measures were not independent. To the extent the samples overlap, the analysis is more an evaluation of "alternative form" reliability."

I would prefer confidence intervals for the correlations rather than p-values.

It would be stronger to assess if patients (clustered within plans) who meet the measure have better outcomes that those that do not.

**Panel Member #6:** *As explained in Q #21, the methods were not adequate.*

**Panel Member #7:** I would favor understanding whether administrative claims and medicare record review elicit similar values of the measure.

**FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction**

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

**ADDITIONAL RECOMMENDATIONS**

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

**Panel Member #1:** I believe the SMP should discuss these concerns before forwarding them to the standing committee.

**Panel Member #2:** Note to NQF staff: we need an interpretative standard for reliability metrics and pearson correlations rather than having each developer cite a standard (or at least cite the authority for the standard).

**Panel Member #3:** On the next round, the measure developers need to include risk adjustment to account for potential difficulty in achieving good BP control in different patient populations.

**Panel Member #4:** *I do not claim expertise in this review and will be eager to learn how others interpret the exclusions (particularly the denominator definition), the use of a binary outcome, and the very high reliability estimates provided.*

**Panel Member #5:** The supporting literature (1b.1) and descriptive performance data (1b.2) are dated. The latter was pre ICD-10.

## 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**

CBP_Evidence_Form_-18-.docx

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

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

Yes.

### 1a. Evidence (subcriterion 1a)

**Measure Number** (*if previously endorsed*)**:** 0018

**Measure Title**: Controlling High Blood Pressure

**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**: 8/1/2019

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

Outcome

☐ Outcome: Click here to name the health outcome

- ☐ 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.

Patients 18-85 years of age with hypertension >>>> Health care providers routinely monitors patients' blood pressure, advises lifestyle modification, and potentially prescribes antihypertensive medication >>>> Patients' blood pressure is adequately controlled (less than 140/90 mm Hg)

>>>> Patients experience fewer cardiovascular events such as stroke and myocardial infarction, as well as death

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

N/A

****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.**

N/A

**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

Table 1. Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults (ACC/AHA)

| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date** | • **Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults**<br>• **ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA November 2017**<br>• **Whelton PK, Carey RM, Aronow WS, Casey DE Jr, Collins KJ, Dennison Himmelfarb C, DePalma SM, Gidding S, Jamerson KA, Jones DW, MacLaughlin EJ, Muntner P, Ovbiagele B, Smith SC Jr, Spencer CC, Stafford RS, Taler SJ, Thomas RJ, Williams KA Sr,** |

| | |
|---|---|
| • **Citation, including page number**<br>• **URL** | **Williamson JD, Wright JT Jr. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension. 2018;71:e13–e115. DOI: 10.1161/ HYP.0000000000000065.**<br>• **URL: https://www.ahajournals.org/doi/pdf/10.1161/HYP.0000000000000065** |
| 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. | Recommendations:<br>BP Goal for Patients With Hypertension<br><br>Recommendation 1: "For adults with confirmed hypertension and known CVD or 10-year ASCVD event risk of 10% or higher, a BP of less than 130/80 mm HG is recommended". *Class I; Level B-R (systolic), Level C-EO (diastolic)*<br><br>Recommendation 2: "For adults with confirmed hypertension, without additional markers of increased CVD risk, a BP target of less than 130/80 mm HG may be reasonable". *Class IIb; Level B-NR (systolic), Level C-EO (diastolic)* |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | The grades assigned by the ACC and AHA to the evidence associated with the recommendation varied by the guideline recommendation. See the question above for the grade given to each guideline recommendation.<br><br>Grade of Evidence:<br>Level B-R (Randomized)<br>• Moderate-quality evidence from 1 or more RCTs<br>• Meta-analyses of moderate-quality RCTs<br>Level B-NR (Nonrandomized)<br>• Moderate-quality evidence from 1 or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies<br>• Meta-analyses of such studies<br>Level C-EO (Expert Opinion)<br>• Consensus of expert opinion based on clinical experience |
| Provide all other grades and definitions from the evidence grading system | Grade of Evidence:<br>Level A<br>• High-quality evidence from more than 1 RCT<br>• Meta-analyses of high-quality RCTs<br>• One or more RCTs corroborated by high-quality registry studies |

| | |
|---|---|
| | Level C-LD (Limited Data)<br>• Randomized or nonrandomized observational or registry studies with limitations of design or execution<br>• Meta-analyses of such studies<br>• Psychological or mechanistic studies in human subjects |
| Grade assigned to the **recommendation** with definition of the grade | The grades assigned by the ACC and AHA to the guideline varied by the guideline recommendation. See the question above for the grade given to each guideline recommendation.<br><br>Grade of Recommendation:<br><br>Class 1 (Strong) Benefit >>> Risk<br>Suggested phrases for writing recommendations:<br>• Is recommended<br>• Is indicated/useful/effective/beneficial<br>• Should be performed/administered/other<br>• Comparative-Effectiveness Phrases †:<br>  ○ Treatment/strategy A is recommended/indicated in preference to treatment B<br>  ○ Treatment A should be chosen over treatment B<br>Class IIb (Weak) Benefit ≥ Risk<br>Suggested phrases for writing recommendations:<br>• May/might be reasonable<br>• May/might be considered<br>• Usefulness/effectiveness is unknown/unclear/uncertain or not well established<br>† For comparative-effectiveness recommendations (COR 1 and IIa; LOE A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated |
| Provide all other grades and definitions from the recommendation grading system | Grade of Recommendation:<br><br>Class IIa (Moderate) Benefit >> Risk<br>Suggested phrases for writing recommendations:<br>• Is reasonable<br>• Can be useful/effective/beneficial<br>• Comparative-Effectiveness Phrases †:<br>  ○ Treatment/strategy A is probably recommended/indicated in preference to treatment B<br>  ○ It is reasonable to choose treatment A over treatment B<br>Class III: No Benefit (Moderate) Benefit = Risk<br>Suggested phrases for writing recommendations: |

| | |
|---|---|
| | • Is not recommended<br>• Is not indicated/useful/effective/beneficial<br>• Should not be performed/administered/other<br>Class III: Harm (Strong) Risk > Benefit<br>Suggested phrases for writing recommendations:<br>• Potentially harmful<br>• Causes harm<br>• Associated with excess morbidity/mortality<br>• Should not be performed/administered/other<br>† For comparative-effectiveness recommendations (COR 1 and IIa; LOE A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | There have been 8 recent meta-analyses addressing the issues of BP reduction and target BP levels for the treatment of hypertension. Although treatment of hypertension was associated with improved outcomes in all 8 meta-analyses, the optimal target BP remains unclear.<br><br>Trials have tested whether more intensive BP control improves major CVD outcomes. Meta-analyses and systematic reviews of these trials provide strong support for the more intensive approach, but the data are less clear in identification of a specific optimal BP target.<br><br>RCTs were included if they met the following eligibility criteria: adults (≥18 years of age) with primary hypertension or hypertension due to CKD; if the intervention included a target BP that was more "intensive" or "lower" than a "standard" or "higher" target BP in the comparator arm; and outcomes included all-cause mortality, cardiovascular mortality, major cardiovascular events, MI, stroke, heart failure, or renal outcomes. Trials were excluded if the primary intent of the study was not specifically to treat or lower BP, were observational studies, or included <100 randomized participants or <400 person years of follow-up, and a minimum of 12 months of follow-up.<br><br>Quantity:<br>A total of 33 publications from 15 studies were considered for inclusion. Publication dates ranged from 1998 through 2015. Target BPs varied for the lower and standard BP groups with 9 studies having a systolic BP target <130 mm Hg for the lower therapy group. Many of the studies included patients with comorbid conditions such as diabetes mellitus, CKD, or were at high risk for cardiovascular disease or progression to end-stage renal disease. Most studies excluded those with prior or recent MI or stroke, secondary hypertension, CHF, or other serious illnesses. |

| | |
|---|---|
| | The below recommendation quality descriptions are the ACC/AHA published systematic review detailing types of studies used in their analysis.<br><br>Recommendation 1:<br>Quality:<br>"Meta-analysis and systematic review of trials that compare more intensive BP reduction to standard BP reduction report that more intense BP lowering significantly reduces the risk of stroke, coronary events, major cardiovascular events, and cardiovascular mortality. In a stratified analysis of these data, achieving an additional 10–mm Hg reduction in SBP reduced CVD risk when compared with an average SBP of 158/82 to 143/76 mm Hg, 144/85 to 137/81 mm Hg, and 134/79 to 125/76 mm Hg. Patients with DM and CKD were included in the analysis. (Specific management details are in Section 9.3 for CKD and Section 9.6 for DM.)"<br><br>Recommendation 2:<br>Quality:<br>"The treatment of patients with hypertension without elevated risk has been systematically understudied because lower-risk groups would require prolonged follow-up to have a sufficient number of clinical events to provide useful information. Although there is clinical trial evidence that both drug and nondrug therapy will interrupt the progressive course of hypertension, there is no trial evidence that this treatment decreases CVD morbidity and mortality. **The clinical trial evidence is strongest for a target BP of 140/90 mm Hg in this population**. However, observational studies suggest that these individuals often have a high lifetime risk and would benefit from BP control earlier in life". |
| Estimates of benefit and consistency across studies | The management and control of high blood pressure to recommended target BP levels has been associated with shift in population BP to lower levels and disease risks.<br><br>In patients with hypertension without elevated cardiovascular risk, there is clinical trial evidence that both drug and nondrug therapy will interrupt the progressive course of hypertension, but there is no trial evidence that this treatment decreases CVD morbidity and mortality. The clinical trial evidence is strongest for a target BP of 140/90 mm Hg in this population.<br><br>The below study descriptions are from excerpts of the ACC/AHA published systematic review detailing the benefits across major studies.<br><br>"Recent trials that address optimal BP targets include SPRINT and ACCORD (Action to Control Cardiovascular Risk in Diabetes), with targets for more intensive (SBP <120 mm Hg) and standard (SBP <140 mm Hg) treatment, and SPS-3, with a more intensive target of <130/80 mm Hg. |

| | |
|---|---|
| | These trials yielded mixed results in achieving their primary endpoints. SPRINT was stopped early, after a median follow-up of 3.26 years, when more intensive treatment resulted in a significant reduction in the primary outcome (a CVD composite) and in all-cause mortality rate.<br><br>In ACCORD, more intensive BP treatment failed to demonstrate a significant reduction in the primary outcome (a CVD composite). However, the incidence of stroke, a component of the primary outcome, was significantly reduced. The standard glycemia subgroup did show significant benefit in ACCORD, and a meta-analysis of the only 2 trials (ACCORD and SPRINT) testing an SBP goal of <120 mm Hg showed significant reduction in CVD events.<br><br>Pooling of the experience from 19 trials (excluding SPRINT) that randomly assigned participants to different BP treatment targets identified a significant reduction in CVD events, MI, and stroke in those assigned to a lower (average achieved SBP/DBP was 133/76 mm Hg) versus a higher BP treatment target. Similar patterns of benefit were reported in 3 other meta-analyses of trials in which participants were randomly assigned to different BP targets and in larger meta-analyses that additionally included trials that compared different intensities of treatment.<br><br>The totality of the available information provides evidence that a lower BP target is generally better than a higher BP target and that some patients will benefit from an SBP treatment goal <120 mm Hg, especially those at high risk of CVD.<br><br>There was agreement across meta-analyses that greater BP lowering appears to be most beneficial for the reduction in risk of major cardiovascular events, MI, stroke, and heart failure. Two studies reported a significant reduction in the risk of all-cause mortality, 3 studies reported reduction in cardiovascular mortality, but no meta-analysis found a significant reduction in the risk of renal events for the lower BP target group compared with a higher BP target group". |
| What harms were identified? | The ACC and AHA didn't provide an explicit discussion of the harms that were discussed in each study supporting the recommendations. However, there was some overall discussion of harms associated with treatment with antihypertensive medications, which largely consist of increased risk of adverse events such as, hypotension, syncope, electrolyte abnormalities, and acute kidney injury. Discussion also pointed to a potential risk of falls and serious fall injuries due to hypotension. Overall, the studies supporting these recommendations found that the benefits of blood pressure-lowering treatments outweigh the harms. |
| Identify any new studies conducted | There have been no new studies that contradict the current body of evidence. |

| | |
|---|---|
| since the SR. Do the new studies change the conclusions from the SR? | |

Table 2. Pharmacologic Treatment of Hypertension in Adults Aged 60 Years or Older to Higher Versus Lower Blood Pressure Targets: A Clinical Practice Guideline From the American College of Physicians and the American Academy of Family Physicians

| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | • **Pharmacologic Treatment of Hypertension in Adults Aged 60 Years or Older to Higher Versus Lower Blood Pressure Targets: A Clinical Practice Guideline From the American College of Physicians and the American Academy of Family Physicians**<br>• **American College of Physicians**<br>• **March 2017**<br>• **Qaseem A, Wilt TJ, Rich R, et al, for the Clinical Guidelines Committee of the American College of Physicians and the Commission on Health of the Public and Science of the American Academy of Family Physicians. Pharmacologic Treatment of Hypertension in Adults Aged 60 Years or Older to Higher Versus Lower Blood Pressure Targets: A Clinical Practice Guideline From the American College of Physicians and the American Academy of Family Physicians. Ann Intern Med. 2017;166:430–437. [Epub ahead of print 17 January 2017]. doi: 10.7326/M16-1785**<br>• **URL: https://annals.org/aim/fullarticle/2598413/pharmacologic-treatment-hypertension-adults-aged-60-years-older-higher-versus** |
| 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. | Recommendations:<br>• Recommendation 1: ACP and AAFP recommend that clinicians initiate treatment in adults aged 60 years or older with systolic blood pressure persistently at or above 150 mm Hg to achieve a target systolic blood pressure of less than 150 mm Hg to reduce the risk for mortality, stroke, and cardiac events. (*Grade: strong Recommendation, high-quality evidence*). ACP and AAFP recommend that clinicians select the treatment goals for adults aged 60 years or older based on a periodic discussion of the benefits and harms of specific blood pressure targets with the patient.<br>• Recommendation 2: ACP and AAFP recommend that clinicians consider initiating or intensifying pharmacologic treatment in adults aged 60 years or older with a history of stroke or transient ischemic attack to achieve a target systolic blood pressure of less than 140 mm Hg to reduce the risk for recurrent stroke. (*Grade: weak recommendation, moderate-quality evidence*). ACP and AAFP recommend that clinicians select the treatment goals for adults |

<table>
<tr><td></td><td>aged 60 years or older based on a periodic discussion of the benefits and harms of specific blood pressure targets with the patient.<br>• Recommendation 3: ACP and AAFP recommend that clinicians consider initiating or intensifying pharmacologic treatment in some adults aged 60 years or older at high cardiovascular risk, based on individualized assessment, to achieve a target systolic blood pressure of less than 140 mm Hg to reduce the risk for stroke or cardiac events. (Grade: weak recommendation, low-quality evidence). ACP and AAFP recommend that clinicians select the treatment goals for adults aged 60 years or older based on a periodic discussion of the benefits and harms of specific blood pressure targets with the patient.</td></tr>
<tr><td>Grade assigned to the evidence associated with the recommendation with the definition of the grade</td><td>The grades assigned by ACP/AAFP to the guideline varied by the guideline recommendation. The grades varied from low-quality to high-quality. See question above for the grade given to each guideline. The table below summarizes the grading of evidence provided in the guideline.<br><br>
<table>
<tr><td></td><td colspan="2">Strength of Recommendation</td></tr>
<tr><td>Quality of Evidence</td><td>Benefits Clearly Outweigh Risks and Burden or Risks and Burden Clearly Outweigh Benefits</td><td>Benefits Finely Balanced with Risks and Burden</td></tr>
<tr><td>High</td><td>Strong</td><td>Weak</td></tr>
<tr><td>Moderate</td><td>Strong</td><td>Weak</td></tr>
<tr><td>Low</td><td>Strong</td><td>Weak</td></tr>
<tr><td colspan="3">Insufficient evidence to determine net benefits or risks</td></tr>
<tr><td colspan="3">* Adopted from the classification developed by the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) workgroup: https://www.ncbi.nlm.nih.gov/pubmed/24404627</td></tr>
</table></td></tr>
<tr><td>Provide all other grades and definitions from the evidence grading system</td><td>No additional grading was provided, grades assigned to the evidence encompass everything within the grading scale.</td></tr>
<tr><td>Grade assigned to the recommendation</td><td>Strength of Recommendation<br>Strong</td></tr>
</table>

| | |
|---|---|
| with definition of the grade | • Benefits clearly outweigh risks and burden, or risks and burden clearly outweigh benefits<br>Weak<br>• Benefits finely balanced with risks and burden |
| Provide all other grades and definitions from the recommendation grading system | No additional grading was provided, grades assigned to the recommendations encompass everything within the grading scale. |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | There is high-quality evidence to show that treatment towards an SBP of less than 150 mm Hg for individuals with a baseline SBP of 160 mm Hg or greater reduces the relative risk of all-cause mortality, absolute risk reduction, stroke and cardiac events.<br><br>There was no statistical significance in reduction of all-cause mortality, absolute risk reduction or cardiac events in studies of lower SBP targets (<140 mm Hg). These studies have low-quality evidence. Lower BP target studies with moderate quality evidence show a reduced risk for stroke and absolute risk reduction compared with higher BP targets. Of these studies, several did not actually achieve the target BP and showed almost no difference between the control and intensive treatment group.<br><br>Lower SBP targets (<140 mm Hg) compared to higher SBP targets (≥140mm Hg) in a subgroup analysis showed similar risk reduction for mortality and cardiac events. The following data show the previously mentioned outcomes. "Mortality RR for target ≥140 mm Hg, 0.91 [CI, 0.84 to 0.99] vs. RR for target <140 mm Hg, 0.84 [CI, 0.74 to 0.95]) and cardiac events (RR for target ≥140 mm Hg, 0.78 [CI, 0.68 to 0.93] vs. RR for target <140 mm Hg, 0.83 [CI, 0.70 to 0.94])".<br><br>Studies that had a target SBP of 140 mmHg or greater had a larger reduction in stoke events than studies that had a target SBP of 140mm Hg. However, it is important to note that there are clinical differences and "significant statistical heterogeneity" which should be taken into consideration when looking at pooled results.<br><br>Below are excerpts that represent the overall quantity and quality of studies examined by the ACP/AAFP.<br><br>Quantity:<br>46 publications representing 21 randomized, controlled trials and 3 cohort studies were included. A total of 15 RCT's were included in the meta-analysis of mortality, stroke, and cardiac events. |

| | |
|---|---|
| | Eight trials compared BP targets, and 13 trials randomly assigned patients to more versus less intensive antihypertensive therapy. Two of the trials included only patients with prior stroke and are considered separately for secondary stroke prevention. Three trials had serious methodological flaws that placed them at high risk of bias, whereas the other 18 trials were judged to have low risk of bias. Because we focused primarily on comparing the effects of more versus less aggressive BP lowering, we conducted sensitivity analyses without 3 trials (2 achieved minimal between-group differences in SBP [≤3 mm Hg], and a third did not report achieved BP) and found similar results. In the following sections on health outcome effects, we present results from the remaining 15 trials.<br><br>Six trials evaluated a total of 41 491 patients and found that treatment targets of SBP less than 140 mm Hg or DBP of 85 mm Hg or lower did not reduce mortality (RR, 0.93 [CI, 0.75 to 1.14]), cardiac events (RR, 0.91 [CI, 0.77 to 1.04]), or stroke. Even though these are large trials with low risk of bias, the evidence should be considered low-strength because the results have important inconsistencies, and because the CIs are relatively wide encompassing the possibility of both marked benefit and harm.<br><br>Recommendation 1:<br>Quality:<br>High-quality evidence showed that treating hypertension in older adults to moderate targets (<150/90 mm Hg) reduces mortality (ARR, 1.64), stroke (ARR, 1.13), and cardiac events (ARR, 1.25). Most benefits apply to such adults regardless of whether they have diabetes. The most consistent and greatest absolute benefit was shown in trials with a higher mean SBP at baseline (>160 mm Hg). Any additional benefit from aggressive BP control is small, with a lower magnitude of benefit and inconsistent results across outcomes.<br><br>Recommendation 2:<br>Quality:<br>Moderate-quality evidence showed that treating hypertension in older adults with previous TIA or stroke to an SBP target of 130 to 140 mm Hg reduces stroke recurrence (ARR, 3.02) compared with treatment to higher targets, with no statistically significant effect on cardiac events or all-cause mortality.<br><br>Recommendation 3:<br>Quality:<br>An SBP target of less than 140 mm Hg is a reasonable goal for some patients with increased cardiovascular risk. The target depends on many factors unique to each patient, including comorbidity, medication burden, risk for adverse events, and cost. Clinicians should individually assess cardiovascular risk for patients. Generally, increased cardiovascular risk includes persons |

| | |
|---|---|
| | with known vascular disease, most patients with diabetes, older persons with chronic kidney disease with estimated glomerular filtration rate less than 45 mL/ min/per 1.73 m2, those with metabolic syndrome (abdominal obesity, hypertension, diabetes, and dyslipidemia), and older persons. For example, among the included studies, SPRINT defined patients with increased cardiovascular risk as those meeting at least 1 of the following criteria: clinical or subclinical cardiovascular disease other than stroke; chronic kidney disease, excluding polycystic kidney disease, with an estimated glomerular filtration rate of 20 to less than 60 mL/min/1.73 m2 of body surface area; 10-year risk for cardiovascular disease of 15% or greater based on the Framingham risk score; or age 75 years or older. This trial found that targeting SBP to less than 120 mm Hg compared with less than 140 mm Hg in adults without diabetes or prior stroke, at high-risk for cardiovascular disease, and with a baseline SBP of less than 140 mm Hg significantly reduced fatal and nonfatal cardiovascular events and all-cause mortality. |
| Estimates of benefit and consistency across studies | The ACP and AAFP found that across all trials, treating high BP in older adults was beneficial. However, most of the evidence came from studies of patients with moderate or severe hypertension (SBP >160 mm Hg) at baseline and, with treatment, achieved SBP targets greater than 140 mm Hg.<br><br>The below excerpts are from the systematic review conducted by the ACP and AAFP.<br><br>In studies with lower SBP targets (<140 mm Hg), low-quality evidence showed no statistically significant reduction in all-cause mortality (RR, 0.93 [CI, 0.75 to 1.14]; ARR, 0.21),cardiac events (RR, 0.91 [CI, 0.77 to 1.04]; ARR, 0.35),or stroke (RR, 0.86 [CI, 0.64 to 0.1.07]; ARR, 0.19) (11–13, 20, 22, 23). For studies with lower BP targets, moderate-quality evidence showed a reduced risk for stroke (RR, 0.79 [CI, 0.59 to 0.99]; ARR, 0.49) compared with higher BP targets (11–13, 20, 22, 23). Many of these studies, however, did not achieve the targeted BP, and there was little difference between the intensive treatment and control groups. Therefore, these studies may not have been able to detect differences in clinical outcomes.<br><br>Most patients aged 60 years or older with an SBP of 150 mm Hg or greater who receive antihypertensive medications will have benefit with acceptable harms and costs from treatment to a BP target of less than 150/90 mm Hg. Although some benefit is achieved by aiming for lower BP targets, most benefit occurs with acceptable harms and costs in the pharmacologic treatment of patients who have an SBP of 150 mm Hg or greater.<br><br>Nine trials provided moderate- to high-strength evidence that BP control to less than 150/90 mm Hg reduces mortality (relative risk [RR], 0.93 [95% CI, 0.85 to 1.00]), cardiac events (RR, 0.83 [CI, 0.71 to 0.96]), and stroke (RR, 0.77 [CI, 0.65 to 0.91]). Six trials overall provide low-strength evidence that |

| | |
|---|---|
| | lower targets (≤140/85) do not reduce mortality (RR, 0.93 [CI, 0.75 to 1.14]), cardiac events (RR, 0.91 [CI, 0.77 to 1.04]), or stroke (RR, 0.86 [CI, 0.64 to 1.07]). However, there were important inconsistencies across these studies, and one large trial showed targeting SBP less than 120 mm Hg in patients at high cardiovascular risk reduced mortality and cardiac events.<br><br>When we removed SPRINT in additional sensitivity analyses, effects on mortality (RR, 0.96 [CI, 0.80 to 1.15]; $I^2$ = 0%) were reduced and effects on cardiac events (RR, 0.88 [CI, 0.74 to 1.04]; $I^2$ = 4.0%) were no longer significant, but effects on stroke remained largely unchanged (RR, 0.74 [CI, 0.56 to 0.99]; $I^2$ = 25.8%). Taken together, SPRINT and the ACCORD trial contribute the most to the uncertainty about the true effect of more intensive BP lowering because of their discrepant results. Both trials compared an SBP target of less than 120 mm Hg versus less than 140 mm Hg in patients with well-controlled hypertension and high cardiovascular risk, but SPRINT found marked reductions in mortality and cardiac events, whereas the ACCORD trial did not. There are several potential reasons that the trials produced different results. The ACCORD trial included only diabetic patients, whereas SPRINT excluded them; the mean age of participants in the ACCORD trial was lower (62 vs. 68 years, though the event rates in both trials were similar); the ACCORD trial was smaller; and SPRINT was stopped early for benefit, which could have exaggerated treatment effects. |
| What harms were identified? | In two studies there were no increased risks for falls, fractures or cognitive impairments for patients who achieved a diastolic blood pressure of less than 70 mm Hg. There was however increased risk for symptomatic hypertension in two trials and in one trial risk for syncope, but this was rated low quality evidence. The patients with these events had very low DBP, SBP, or both.<br><br>Regardless of treatment to higher or lower BP, low quality evidence showed no difference in renal outcomes, functionals status or risk for falls. Moderate quality evidence showed no difference in cognitive decline, dementia, fractures or quality of life for treatment to higher versus lower BP.<br><br>The ACP/AAFP did not specifically examine nonpharmacologic options for reducing BP, but acknowledges that there are typically fewer side-effects associated with that treatment as compared to pharmacologic treatments for hypertension.<br><br>Overall, there were mixed results for associated adverse events in the included studies. For studies treating to lower BP, 4 out of 10 trials had withdrawals due to adverse events. The most commonly reported adverse events were cough and hypotension. |

| | |
|---|---|
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | There have been no new studies that contradict the current body of evidence. |

**1a.4 OTHER SOURCE OF EVIDENCE**

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

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

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.*

One out of every three Americans have hypertension, or high blood pressure. Even with the availability of effective treatment options, only about half (54%) of these people have their high blood pressure under control (Merai et al, 2016). Improvements in quality or better control of blood pressure as related to this measure would help significantly reduce the probability of serious and costly complications, including coronary artery disease, congestive heart failure, stroke, ruptured aortic aneurysm, renal disease and retinopathy.

Merai R, Siegel C, Rakotz M, Basch P, Wright J, Wong B; DHSc., Thorpe P. CDC Grand Rounds: A Public Health Approach to Detect and Control Hypertension. MMWR Morb Mortal Wkly Rep. 2016 Nov 18;65(45):1261-1264.

**1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile*

*range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

The following data are extracted from HEDIS data collection and reflect the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by the mean, standard deviation, minimum health plan performance, maximum health plan performance, performance percentiles (10th, 25th, 50th, 75th, and 90th percentile) and the interquartile range. Data is stratified by year and product line (i.e. commercial, Medicare, Medicaid) at the health plan level.

Controlling High Blood Pressure

N = Number of Health Plans

YEAR = Measurement Year

Commercial

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

2016|352|59%|12%|22%|44%|50%|57%|68%|75%|90%|18%

2017|350|58%|12%|26%|44%|51%|58%|66%|75%|92%|15%

2018|403|55%|21%|0%|9%|52%|60%|67%|74%|84%|15%

Medicaid

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

2016|259|56%|12%|25%|40%|48%|57%|65%|72%|90%|17%

2017|264|57%|13%|2%|42%|49%|59%|66%|71%|85%|17%

2018|248|59%|13%|0%|46%|53%|61%|67%|72%|85%|14%

Medicare

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

2016|466|70%|13%|24%|51%|61%|71%|81%|84%|97%|19%

2017|460|71%|14%|8%|54%|63%|74%|80%|86%|96%|18%

2018|483|69%|11%|0%|57%|64%|71%|76%|81%|100%|12%

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

N/A

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

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

The report described racial and ethnic disparities among beneficiaries 18 and older with a diagnosis of hypertension who had their blood pressure under control. Approximately 69% of White and Asian or Pacific Islander beneficiaries who had a diagnosis of hypertension had their blood pressure under control while 59% of Blacks had their blood pressure under control. Hispanics underperformed compared to Whites and Asian or Pacific Islanders, but by less than 3%. About 67% of Hispanics had their blood pressure under control.

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

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

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

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

Although HEDIS measures are not stratified by race and ethnicity, there is an abundance of disparities data related to hypertension.

The prevalence of hypertension rises with increasing age and varies by race. Data from the Framingham Heart Study found that among adults with a baseline SBP/DBP of 130 to 139/85 to 89 mm Hg, 49.5 percent of adults 65 to 94 years of age developed hypertension compared to 37.3 percent of adults 35 to 64 years of age (Vasan, 2001). Among various races, blacks have the highest prevalence of hypertension across the world (Benjamin et al, 2019). Between 2013-2016 the prevalence of hypertension for adults =20, among non-Hispanic black males and females was 58.6 percent and 56.0 percent, among non-Hispanic white males and females was 48.2 percent and 41.3 percent, among non-Hispanic Asian males and females was 46.4 percent and 36.4 percent, and among Hispanic males and females was 47.4 and 40.8 percent, respectively (Benjamin et al, 2019).

There are disparities in awareness, treatment, and control of hypertension. Data between 2013-2016 from NHANES showed that those with hypertension who were =20 years of age, 64.7 percent were aware of their condition, 53.4 percent were under current treatment, and 24.7 percent had their hypertension under control (Benjamin et al, 2019). When comparing races, non-Hispanic whites and blacks are more aware of their hypertension than Hispanic and non-Hispanic Asian adults (Benjamin et al, 2019). Based on Medicare Advantage data in 2017, the percent of adequately controlled blood pressure for hypertensive adults 18-85, among Asian or Pacific Islander beneficiaries was 69.4 percent, among non-Hispanic white beneficiaries was 68.6 percent, among Hispanic beneficiaries was 66.7 percent, among non-Hispanic black beneficiaries was 58.9 percent (CMS, 2019). Older adults are more likely to be aware and receive treatment for their hypertension. An average of 54.4 percent of adults 40 and older had their hypertension in control compared to 10.2 percent of adults between 20-39 years of age (Benjamin et al, 2019).

Benjamin EJ et al., Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association. Circulation. 2019;139:e56–e528. DOI: 10.1161/CIR.0000000000000659

Vasan RS, Larson MG, Leip EP, Kannel WB, Levy D. Assessment of frequency of progression to hypertension in non-hypertensive participants in the Framingham Heart Study: a cohort study. Lancet. 2001;358:1682–1686. doi: 10.1016/S0140-6736(01)06710-1.

## 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):*

Cardiovascular, Cardiovascular : Hypertension, Endocrine : Diabetes

**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:** 0018_CBP_Value_Sets_Fall_2019-637002741932672877.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.

Yes

**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.

This measure has undergone several changes since its last maintenance review. The measure now uses a purely administrative approach to identify patients for the denominator, assesses a blood pressure threshold of <140/90 mm Hg for all patients and incorporates the use of blood pressure readings from certain remote

monitoring devices. In addition to that, there have been minor changes to the value sets and medication lists to reflect current practice.

NCQA added a hospice exclusion to most HEDIS measures in 2016. The focus of hospice care is not to cure illnesses of patients, but rather to improve comfort and quality of life for those with less than six months to live. Most HEDIS quality measures are focused on health screenings or treatments that are not clinically appropriate or beneficial for those who are at end of life. Many of these screenings and treatments would also be uncomfortable for hospice patients, add undue burden and have no impact on improving length or quality of life. Therefore, including individuals who are receiving hospice in our HEDIS quality measures is inappropriate.

In addition, NCQA added exclusion criteria for adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings. We recognize that for individuals with limited life expectancy, advanced illness or more complex clinical situations, the focus of this measure may not be relevant or in line with the patient's goals of care. By implementing this set of exclusions, those providing care to the frail and advanced illness population can focus on care that's more appropriate for their conditions and health status. Attention can be more focused on quality measures that capture services and care processes that are most relevant for this population (e.g., improving care transitions, getting follow-up after acute care episodes, or avoiding preventable hospitalizations).This measure has undergone several changes since its last maintenance review. The measure now uses a purely administrative approach to identify patients for the denominator, assesses a blood pressure threshold of <140/90 mm Hg for all patients and incorporates the use of blood pressure readings from certain remote monitoring devices. In addition to that, there have been minor changes to the value sets and medication lists to reflect current practice.

NCQA added a hospice exclusion to most HEDIS measures in 2016. The focus of hospice care is not to cure illnesses of patients, but rather to improve comfort and quality of life for those with less than six months to live. Most HEDIS quality measures are focused on health screenings or treatments that are not clinically appropriate or beneficial for those who are at end of life. Many of these screenings and treatments would also be uncomfortable for hospice patients, add undue burden and have no impact on improving length or quality of life. Therefore, including individuals who are receiving hospice in our HEDIS quality measures is inappropriate.

In addition, NCQA added exclusion criteria for adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings. We recognize that for individuals with limited life expectancy, advanced illness or more complex clinical situations, the focus of this measure may not be relevant or in line with the patient's goals of care. By implementing this set of exclusions, those providing care to the frail and advanced illness population can focus on care that's more appropriate for their conditions and health status. Attention can be more focused on quality measures that capture services and care processes that are most relevant for this population (e.g., improving care transitions, getting follow-up after acute care episodes, or avoiding preventable hospitalizations).

**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).*

Patients whose most recent blood pressure level was <140/90 mm Hg during the measurement year.

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

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

There are two data sources and approaches used for collecting data reporting the numerator for this measure: Administrative Claims and Medical Record Review

ADMINISTRATIVE CLAIMS

Use codes (See code value sets located in question S.2b.) to identify the most recent BP reading taken during an outpatient visit, a nonacute inpatient encounter, or remote monitoring event during the measurement year.

The blood pressure reading must occur on or after the date when the second diagnosis of hypertension (identified using the event/diagnosis criteria).

The patient is numerator compliant if the blood pressure is <140/90 mm Hg. The patient is not compliant if the blood pressure is >=140/90 mm Hg, if there is no blood pressure reading during the measurement year or if the reading is incomplete (e.g., the systolic or diastolic level is missing). If there are multiple blood pressure readings on the same date of service, use the lowest systolic and lowest diastolic blood pressure on that date as the presentative blood pressure.

Organizations that use CPT Category II codes to identify numerator compliance for this indicator must search for all codes in the following value sets and use the most recent codes during the measurement year to determine numerator compliance for both systolic and diastolic levels.

VALUE SET / NUMERATOR COMPLIANCE

Systolic Less Than 140 Value Set / Systolic compliant

Systolic Greater Than or Equal to 140 Value Set / Systolic not compliant

Diastolic Less Than 80 Value Set / Diastolic compliant

Diastolic 80-89 Value Set / Diastolic compliant

Diastolic Greater Than or Equal to 90 Value Set / Diastolic not compliant

See attached code value sets.

MEDICAL RECORD REVIEW

The number of patients in the denominator whose most recent blood pressure (both systolic and diastolic) is adequately controlled during the measurement year. For a patient's blood pressure to be controlled the systolic and diastolic blood pressure must be <140/90 mm hg (adequate control). To determine if a member's blood pressure is adequately controlled, the representative blood pressure must be identified.

All eligible blood pressure measurements recorded in the record must be considered. If an organization cannot find the medical record, the patient remains in the measure denominator and is considered noncompliant for the numerator.

Use the following guidance to find the appropriate medical record to review.

- Identify the patient's PCP.

- If the patient had more than one PCP for the time-period, identify the PCP who most recently provided care to the patient.

- If the patient did not visit a PCP for the time-period or does not have a PCP, identify the practitioner who most recently provided care to the patient.

- If a practitioner other than the patient's PCP manages the hypertension, the organization may use the medical record of that practitioner.

Identify the most recent blood pressure reading noted during the measurement year.

The blood pressure reading must occur on or after the date when the second diagnosis of hypertension (identified using the event/diagnosis criteria) occurred.

Do not include BP readings:

- Taken during an acute inpatient stay or an ED visit.

- Taken on the same day as a diagnostic test or diagnostic or therapeutic procedure that requires a change in diet or change in medication on or one day before the day of the test or procedure, with the exception of fasting blood tests.

- Reported by or taken by the patient.

BP readings from remote monitoring devices that are digitally stored and transmitted to the provider may be included. There must be documentation in the medical record that clearly states the reading was taken by an electronic device, and results were digitally stored and transmitted to the provider and interpreted by the provider.

Identify the lowest systolic and lowest diastolic BP reading from the most recent BP notation in the medical record. If multiple readings were recorded for a single date, use the lowest systolic and lowest diastolic BP on that date as the representative BP. The systolic and diastolic results do not need to be from the same reading.

The patient is not compliant if the BP reading is =140/90 mm Hg or is missing, or if there is no BP reading during the measurement year or if the reading is incomplete (e.g., the systolic or diastolic level is missing).

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

Patients 18-85 years of age who had at least two visits on different dates of service with a diagnosis of hypertension during the measurement year or the year prior to the measurement year.

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

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

Patients who had continuous enrollment in the measurement year. No more than one gap in continuous enrollment of up to 45 days during the measurement year. If the patient has Medicaid, then no more than a 1-month gap in coverage.

Patients are identified for the denominator using claim/encounter data.

Patients who had at least two visits on different dates of service with a diagnosis of hypertension during the measurement year or the year prior to the measurement year. Visit type need not be the same for the two visits.

Any of the following combinations meet criteria:

- Outpatient visit with any diagnosis of hypertension

- A telephone visit with any diagnosis of hypertension

- An online assessment with any diagnosis of hypertension

Only one of the two visits may be a telephone visit, an online assessment or an outpatient telehealth visit. Identify outpatient telehealth visits by the presence of a telehealth modifier or the presence of a telehealth POS code associated with the outpatient visit.

See attached code value sets.

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

This measure excludes adults in hospice. It also excludes adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings.

Additionally, this measure excludes patients with evidence of end-stage renal disease, dialysis, nephrectomy, or kidney transplant on or prior to the December 31 of the measurement year. It also excludes female patients with a diagnosis of pregnancy during the measurement year, and patients who had a nonacute inpatient admission during the measurement year.

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

ADMINISTRATIVE CLAIMS

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

Exclude adults who meet any of the following criteria:

- Medicare members 66 years of age and older as of December 31 of the measurement year who meet either of the following:

-- Enrolled in an Institutional SNP (I-SNP) any time during the measurement year.

-- Living long-term in an institution any time during the measurement year as identified by the LTI flag in the Monthly Membership Detail Data File. Use the run data of the file to determine if a patient had an LTI flag during the measurement year.

- Members 66-80 years of age as of December 31 of the measurement year (all product lines) with frailty and advanced illness. Patients must meet BOTH of the following frailty and advanced illness criteria to be excluded:

1. At least one claim/encounter for frailty during the measurement year.

2. Any of the following during the measurement year or the year prior to the measurement year (count services that occur over both years):

-- At least two outpatient visits, observation visits, ED visits, nonacute inpatient encounters or nonacute inpatient discharges (instructions below) on different dates of service, with an advanced illness diagnosis. Visit type need not be the same for the two visits. To identify a nonacute inpatient discharge:

1. Identify all acute and nonacute inpatient stays.

2. Confirm the stay was for nonacute care based on the presence of a nonacute code on the claim.

3. Identify the discharge date for the stay.

-- At least one acute inpatient encounter with an advanced illness diagnosis.

-- At least one acute inpatient discharge with an advanced illness diagnosis. To identify an acute inpatient discharge:

1. Identify all acute and nonacute inpatient stays.

2. Exclude nonacute inpatient stays.

3. Identify the discharge date for the stay.

-- A dispensed dementia medication.

DEMENTIA MEDICATIONS

DESCRIPTION / PRESCRIPTION

Cholinesterase inhibitors / Donepezil; Galantamine; Rivastigmine

Miscellaneous central nervous system agents / Memantine

- Members 81 years of age and older as of December 31 of the measurement year (all product lines) with frailty during the measurement year.

Exclude patients with evidence of end-stage renal disease, dialysis, nephrectomy, or kidney transplant on or prior to December 31 of the measurement year, female patients with a diagnosis of pregnancy during the measurement year, and patients who had a nonacute inpatient admission during the measurement year. To identify nonacute inpatient admissions:

1. Identify all acute and nonacute inpatient stays.

2. Confirm the stay was for nonacute care based on the presence of a nonacute code on the claim.

3. Identify the admission date for the stay.

See attached code value sets.

MEDICAL RECORD REVIEW

Exclusionary evidence in the medical record must include a note indicating diagnosis of pregnancy or evidence of a nonacute inpatient admission during the measurement year, or evidence of ESRD, dialysis, nephrectomy or kidney transplant any time during the patient's history through December 31 of the measurement year.

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

N/A

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

No risk adjustment or risk stratification

If other:

**S.12. Type of score:**

Rate/proportion

If other:

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

Better quality = Higher score

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

STEP 1: Determine the eligible population. To do so, identify adults who meet all specified criteria.

- AGES: 18-75 years as of December 31 of the measurement year.

- EVENT/DIAGNOSIS: Identify patients with hypertension in two ways: by claim/encounter data and by medical record data. SEE responses in S.6 and S.7 for eligible population and denominator criteria and details.

STEP 2: Exclude patients who meet the exclusion criteria. SEE responses in S.8 and S.9 for denominator exclusion criteria and details.

STEP 3: Determine the number of patients in the eligible population who had a blood pressure reading during the measurement year through the search of administrative data systems or medical record data.

STEP 4: Identify the lowest systolic and lowest diastolic blood pressure reading from the most recent blood pressure notation in the medical record.

STEP 5: Determine whether the result was <140/90 mm Hg.

STEP 6: Calculate the rate by dividing the numerator (STEP 5) by the denominator (after exclusions) (STEP 2).

**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.

Plans may report this measure using a systematic sample of 411 members. Plans are instructed to list and sort all eligible members for a measure. NCQA then provides plans with a Random Number Table that is released towards the end of the measurement year. The Random Number table lists a value that is used to determine which members from the eligible populations (i.e., every nth member) for whom numerator compliance will be determined.

**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, Electronic Health Data, Electronic Health Records, Paper Medical Records

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

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

This measure is based on administrative claims and medical record documentation collected in the course of providing care to health plan patients. NCQA collects Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from health plans via NCQA's online data submission system.

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

No data collection instrument provided

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

Health Plan

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

Outpatient Services

If other:

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

N/A

**2. Validity – See attached Measure Testing Submission Form**

CBP_Testing_Form_-18-.docx

**2.1 For maintenance of endorsement**

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

Yes

**2.2 For maintenance of endorsement**

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

Yes

**2.3 For maintenance of endorsement**

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

No - This measure is not risk-adjusted.

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

**Measure Number** (*if previously endorsed*)**:** 0018
**Measure Title**: Controlling High Blood Pressure
**Date of Submission**: 8/1/2019

**Type of Measure:**

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

**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: Click here to describe | ☐ other: Click here to describe |

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

N/A

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

Testing of performance measure score with beta binomial reliability and testing of construct validity with the Pearson Correlation were performed using HEDIS 2019 plan level data, measurement year 2018.

**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*)

This measure assesses whether adults enrolled in commercial, Medicare, and Medicaid plans who had a diagnosis of hypertension (HTN) had their blood pressure adequately controlled under 140/90 mm Hg. Therefore, testing was done at the health-plan level, which is appropriate for the level of reporting for this measure.

We calculated the measure score reliability and construct validity from HEDIS data that included 403 commercial plans, 248 Medicaid plans, and 483 Medicare plans. The sample included all commercial, Medicare, and Medicaid health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

**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*)

Table 1 below provides a description of the data submitted for 2018, including the median denominator size per plan. Data are summarized at the health plan level and stratified by plan type (i.e. commercial, Medicaid, Medicare). Since data can be collected and reported from two data sources (administrative claims and medical record review), the vast majority of plans use a combination of data from administrative claims data and a sample of 411 of medical records they review to report their performance rates.

Table 1. Median denominator size per plan for Controlling High Blood Pressure, 2018.

| Product Line | Number of Plans | Median Denominator Size/Plan |
|---|---|---|
| Commercial | 403 | 411 |
| Medicaid | 248 | 411 |
| Medicare | 483 | 411 |

**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**.

Reliability:

Reliability of the health plan measure score was tested using a beta-binomial calculation. This analysis included the entire HEDIS data sample (described above).

Validity:

Validity of the health plan measure was demonstrated through construct validity using the entire HEDIS data sample (described above).

**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 did not analyze social risk factors. This measure of health plan performance is specified to be reported separately by commercial, Medicaid, and Medicare plan types, which serves as a proxy for income and other socioeconomic factors. SEE 2b3.2 for further discussion on research on impact of social risk on this measure.

---

**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*)

Reliability was estimated by using the Beta-binomial model (Adams, 2009) for this health plan measure. Beta-binomial is appropriate for estimating the reliability of pass/fail rate measures. Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

**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*)
Table 2 provides the reliability for the overall measure as shown by the Beta-binomial model as well as the distribution of individual plan reliability.

Table 2. Overall Beta-binomial statistic and distribution of plan reliability for commercial, Medicaid, and Medicare product lines, 2018

| Product Line | Overall Reliability | Min | Percentiles | | | | | Max |
|---|---|---|---|---|---|---|---|---|
| | | | 10th | 25th | 50th | 75th | 90th | |
| Commercial | 0.999 | 0.932 | 0.990 | 0.991 | 0.991 | 0.992 | 1.000 | 1.000 |
| Medicaid | 0.982 | 0.568 | 0.939 | 0.950 | 0.995 | 0.961 | 0.966 | 1.000 |
| Medicare | 0.985 | 0.862 | 0.974 | 0.975 | 0.976 | 0.978 | 0.980 | 1.000 |

**2a2.4 What is your interpretation of the results in terms of demonstrating reliability**? (i.*e., what do the results mean and what are the norms for the test conducted?*)
The values for the beta-binomial statistic across all product lines for the health plan level measure are all greater than 0.7, indicating the measure has very good reliability. The 10-90th percentile distribution of health plan level-reliability on this measure show the vast majority of health plans not only exceeded the minimally accepted threshold of 0.7 but also exceeded 0.9. Strong reliability is demonstrated since the majority of variance is due to signal and not to noise.

---

**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 <u>performance measure score</u> 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)*
We tested for construct validity of the Controlling Blood Pressure measure by exploring whether it was correlated with two similar measures of quality which is described below.

- Comprehensive Diabetes Care: HbA1c Control (< 8%): The percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent HbA1c level is < 8.0% during the measurement year.

- Comprehensive Diabetes Care: HbA1c Poor Control (> 9%): The percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent HbA1c level is > 9.0% during the measurement year.

These measures were chosen for construct validity testing because they are similarly focused on the management of a chronic condition but aimed at different biological markers. We hypothesized that a plan that does well on one measure focused on the management of blood pressure for patients with hypertension will likely do well on other measures focused on the management of other chronic conditions, such as blood glucose for patients with diabetes. Note: The HbA1c Poor Control measure is a "lower is better quality" measure. This means that plans that are performing well will have low rates on this measure.

To test this correlation, we used a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable. Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

** Note: All HEDIS value sets are updated annually with the most current codes available. The information below details the process we used to convert value sets that used ICD-9 codes to ICD-10 codes in 2015. **

ICD-10 CONVERSION:

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

Steps in ICD-9 to ICD-10 Conversion Process

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

Tools Used to Identify/Map to ICD-10

All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).

GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

*Expert Participation*

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

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

The results from construct validity testing of the health plan level measure are presented by product line in Tables 3a, 3b, and 3c below.

Table 3a. Correlations between CBP and CDC HbA1c measures in Commercial Health Plans, 2018.

| | **Pearson Correlation Coefficients** | |
|---|---|---|
| | CDC – HbA1c Control | CDC – HbA1c Poor Control |
| CBP | 0.810 | -0.824 |

Note: All correlations are significant at p<0.0001

Table 3b. Correlations between CBP and CDC HbA1c measures in Medicare Health Plans, 2018.

| | **Pearson Correlation Coefficients** | |
|---|---|---|
| | CDC – HbA1c Control | CDC – HbA1c Poor Control |
| CBP | 0.519 | -0.577 |

Note: All correlations are significant at p<0.0001

Table 3c. Correlations between CBP and CDC HbA1c measures in Medicaid Health Plans, 2018.

| | **Pearson Correlation Coefficients** | |
|---|---|---|
| | CDC – HbA1c Control | CDC – HbA1c Poor Control |
| CBP | 0.795 | -0.820 |

Note: All correlations are significant at p<0.0001

**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?*)

Across all product lines, these correlations are moderate to very strong and statistically significant.

---

**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*)

We did not perform testing of the following exclusions for this submission:

- End-stage renal disease (ESRD)
- Dialysis
- Nephrectomy
- Kidney Transplant
- Pregnancy
- Nonacute inpatient admission

NCQA engaged expert panels to inform the face validity of these exclusions for this measure, which aligns with evidence focused on the general population of people with hypertension. This measure has been reviewed by NCQA's Diabetes Measurement Advisory Panel, Cardiovascular Measurement Advisory Panel, Technical Measurement Advisory Panel, and the Committee on Performance Measurement. The measure also received public comment feedback upon initial development.

Hospice, I-SNPs and Long-Term Care Institutions

These exclusions were also not formally tested for this submission. This measure is designed to be scientifically valid and feasible for comparing the quality of care provided to general populations, such as healthy older adults or those with a single condition. Patients receiving hospice, enrolled in an I-SNP, or residing in a long-term care institution would likely have different care needs and quality concerns, therefore they are excluded from this measure.

Advanced Illness and Frailty

For HEDIS 2019 (measurement year 2018), NCQA added exclusions for advanced illness and frailty to the Controlling Blood Pressure measure. NCQA decided to explore implementing these exclusions, recognizing that for individuals with limited life expectancy, advanced illness and frailty, the focus of this measure may not be clinically appropriate, relevant or in line with the patient's goals of care. We performed a review of literature on different approaches to defining advanced illness and used this, along with feedback received from expert work groups, measurement advisory panels and public comment to create a list of illnesses, conditions and service codes to be included in testing. The conditions included: dementia and other neurodegenerative conditions, emphysema, end stage renal disease (ESRD), heart failure, liver failure, metastatic cancer, pulmonary fibrosis and respiratory failure.

NCQA then conducted a search of ICD-10 codes that were relevant to each of the conditions to create value sets for testing. To identify those with dementia, NCQA also included drug codes for medications such as

donepezil hydrochloride and galantamine hydrobromide, to capture those who may not carry a diagnosis of dementia but are prescribed a drug for treatment.

The proxy for frailty was developed based on previously studied approaches[1, 2, 3] and feedback received from expert work groups and measurement advisory panels. The proxy is comprised of HCPCS and ICD-10 codes for diagnoses or services that can indicate when an individual is frail or dependent in activities of daily living. Examples include: gait abnormality, abnormal loss of weight and underweight, adult failure to thrive, debility, fall, pressure ulcer, durable medical equipment (hospital bed, walker, portable or home oxygen, wheelchair), bed confinement, palliative care and age-related physical debility. Members met the frailty proxy criteria if they had a claim for any of the codes included in the frailty code set in the measurement year.

To determine the feasibility and impact of applying these exclusions to the measure, NCQA used a research database that consisted of two years of inpatient, outpatient, and pharmacy claims for members age 18 and older enrolled in a sample of Medicare Advantage plans (N=19). NCQA compared several approaches for identifying the advanced illness and frailty populations, examining different age ranges and diagnosis positions and their impact on the denominator size. The results of those queries along with input from the expert work groups, measurement advisory panels and public comment led us to determine that the best approach for identifying the advanced illness and frailty population that should be excluded from the measure was to apply the following criteria:

- Adults 66–80 years of age as of December 31 of the measurement year (all product lines) with frailty and advanced illness
- Adults 81 years of age and older as of December 31 of the measurement year with frailty any time on or between July 1 of the year prior to the measurement year and the end of the measurement year.

**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*)
Table 4 shows the results of applying the advanced illness and frailty exclusions to the Controlling High Blood Pressure measure.

Table 4. Impact of applying the exclusions for advanced illness and frailty

| Number of Plans (N) | Average Number Excluded | Average % Removed by Exclusion |
|---|---|---|
| 26 | 784 | 3.8 |

[1] Faurot, K.R., Funk, M.J., Pate, V., Brookhart, M.A., Patrick, A., Hanson, L.C., Castillo, W.C., Stürmer, T. 2015. Using Claims Data to Predict Dependency in Activities of Daily Living as a Proxy for Frailty. Pharmacoepidemiology and Drug Safety. 24(1): 59-66.

[2] Segal, J.B., Chang, H.Y., Du, Y., Walston, J.D., Carlson, M.C., Varadhan, R. 2017. Development of a

Claims-Based Frailty Indicator Anchored to a Well-Established Frailty Phenotype. Medical Care. 55(7): 716-722.

[3] Davidoff A.J., A. Hurrida, I.H. Zuckerman, S.M. Lichtman, N. Pandya, A. Hussain, F. Hendrick, J.P. Weiner, X. Ke, M.J. Edelman. 2013. A Novel Approach to Improve Health Status Measurement in Observational Claims-Based Studies of Cancer Treatment and Outcomes. J Geriatr Oncol. 4(2):157–165.

**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*)

Advanced Illness and Frailty

The advanced illness and frailty exclusion had a small impact on the eligible population: 3.8% on average were removed for advance illness and frailty. Feedback from NCQA's expert work groups and measurement advisory panels, as well as public comment feedback, supported the application of these exclusions to the Controlling High Blood Pressure measure for clinical reasons. By implementing these exclusions, those providing care to patients with advanced illness and frailty can focus on care that is more appropriate for their conditions and health status. Attention can be more focused on quality measures that capture services and care processes that are most relevant for this population (e.g., improving care transitions, getting follow-up after acute care episodes, or avoiding preventable hospitalizations).

---

**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** Click here to enter number of factors **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.**

N/A

**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**.

NCQA recognizes that there is a growing body of literature that might support risk adjustment or stratification of intermediate outcome measures. However, at this time, NCQA does not currently risk adjust this measure given the potential to mask poor performance and disparities in care.

NCQA conducted a study on the Controlling High Blood Pressure measure among Medicare Advantage plans to assess whether to account for a member's socioeconomic status (SES) when comparing plan performance. A qualitative assessment included key informant interviews exploring ways in which SES may affect performance on this and other select HEDIS measures, and whether there was a conceptual basis for case-mix adjustment or other strategies. In the quantitative analysis, we assessed whether SES affected plan performance, using member low-income status, dual eligibility, and disability as proxies for SES. For this measure, adjusting for SES did not have a meaningful impact on results. When adjusting for disparity in performance between low- and high-SES populations, plan ranks were not substantially impacted. When accounting for clinical and demographic factors, we found that low-SES beneficiaries were as likely, or more likely, to receive recommended care as high-SES beneficiaries. Our results suggest there is neither a conceptual nor empirical basis for risk adjustment for this measure.

**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?
N/A

**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.**
N/A

**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*)
N/A

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

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

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

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

**2b3.9. Results of Risk Stratification Analysis**:

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

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

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

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

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure.

To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p-value of the test statistic is less than 0.05, then the two plans' performance is significantly different from each other.

**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*)

Table 4. Variation in Performance for Commercial, Medicaid, and Medicare health plans, 2018.

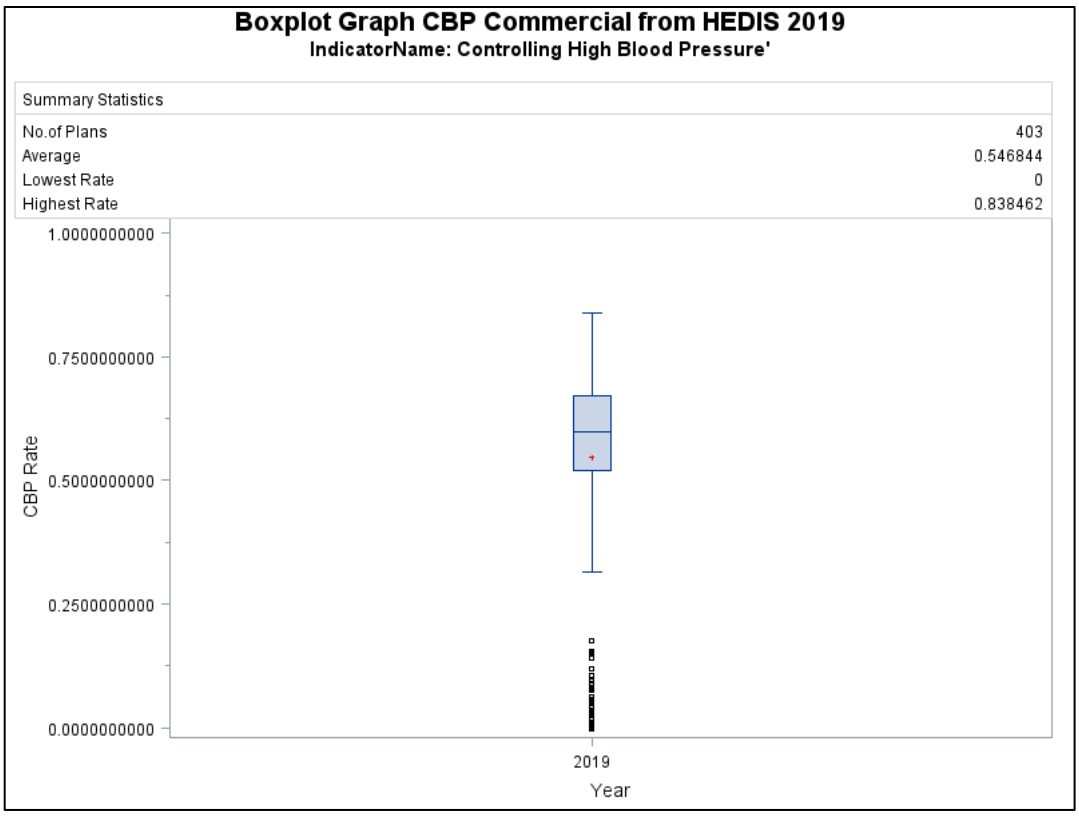
| Plan Type | N | Avg. | St Dev (%) | 10th (%) | 25th (%) | 50th (%) | 75th (%) | 90th (%) | IQR (%) | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| Commercial | 403 | 54.68 | 20.93 | 9.15 | 52.07 | 59.85 | 67.15 | 73.72 | 15.09 | <0.0001 |
| Medicaid | 248 | 58.86 | 12.54 | 45.50 | 52.68 | 60.92 | 66.91 | 72.26 | 14.23 | <0.0001 |
| Medicare | 483 | 69.49 | 10.56 | 57.18 | 64.23 | 70.80 | 75.91 | 80.78 | 11.68 | <0.0001 |

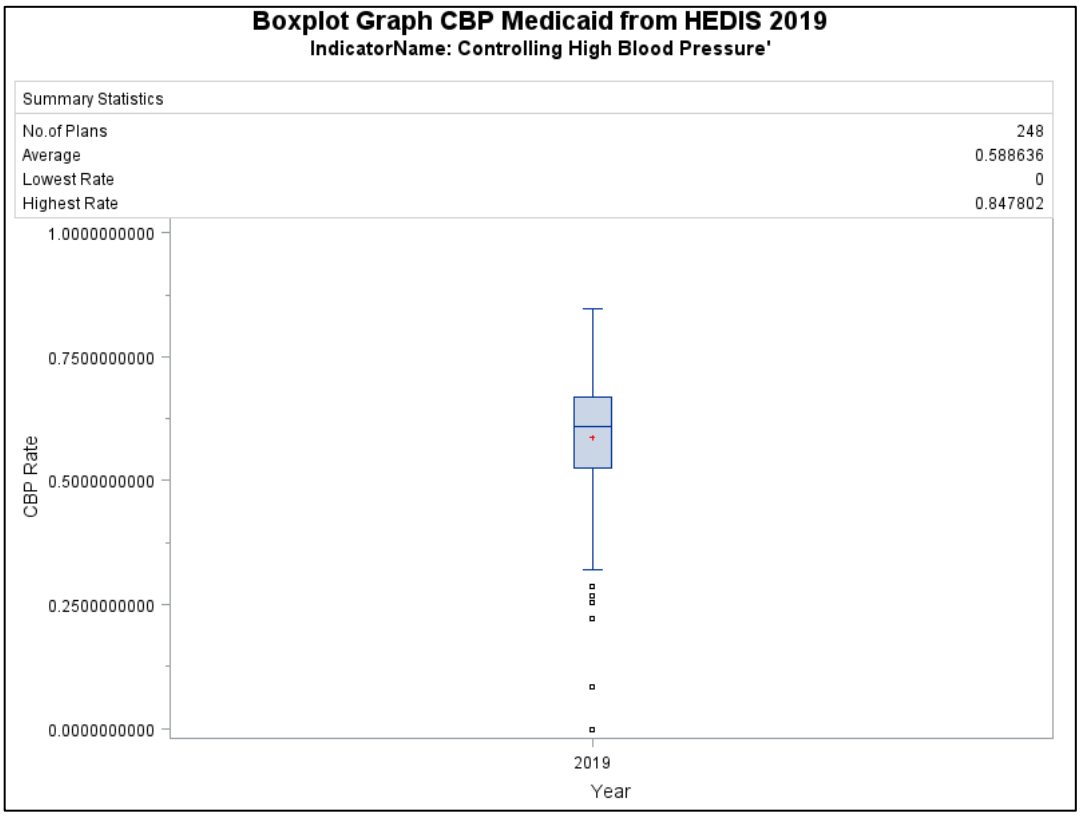
N = Number of plans reporting

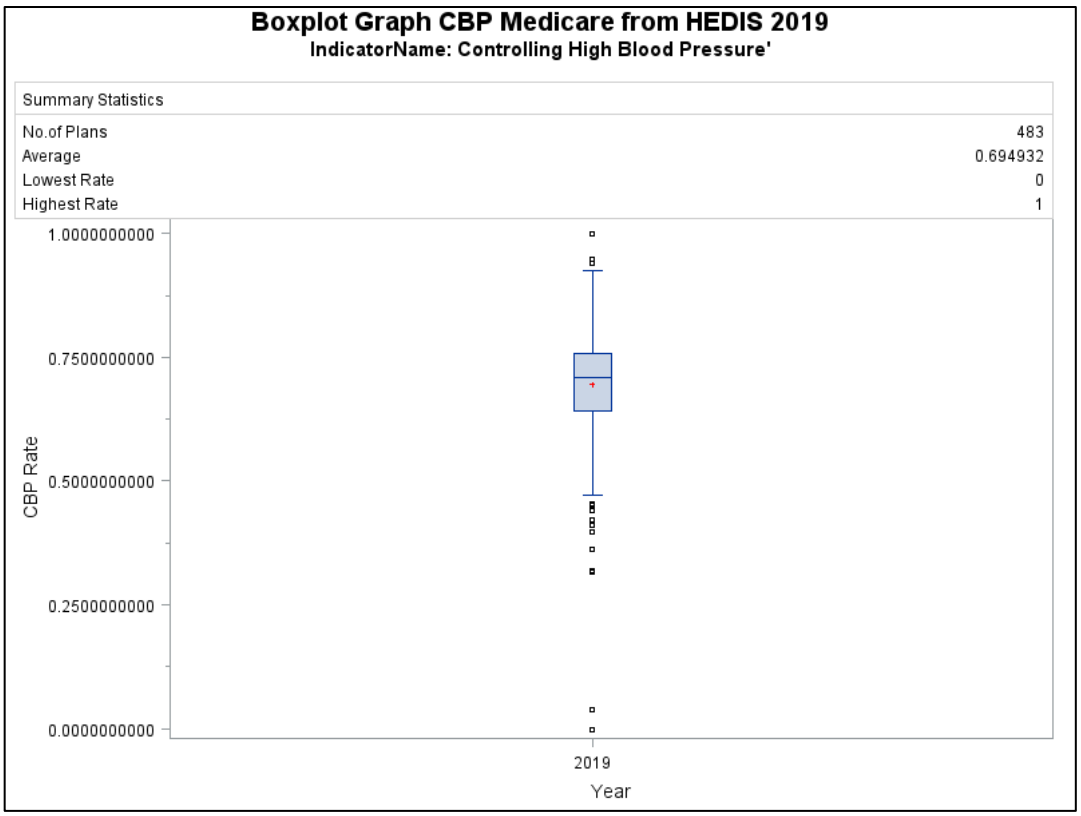
IQR = Interquartile range

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

Box plots for HEDIS 2019 (Measurement year 2018) Variation in Performance Across Health Plans are included below for your reference.


Boxplot Graph CBP Commercial from HEDIS 2019
IndicatorName: Controlling High Blood Pressure'
Summary Statistics
No.of Plans 403
Average 0.546844
Lowest Rate 0
Highest Rate 0.838462
1.0000000000
0.7500000000
0.5000000000
0.2500000000
0.0000000000
CBP Rate
2019
Year


Boxplot Graph CBP Medicaid from HEDIS 2019
IndicatorName: Controlling High Blood Pressure'
Summary Statistics
No.of Plans 248
Average 0.588636
Lowest Rate 0
Highest Rate 0.847802
1.0000000000
0.7500000000
0.5000000000
0.2500000000
0.0000000000
CBP Rate
2019
Year




**2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities?** (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
The results above indicate there is meaningful difference in performance. Across all product lines, the difference between the 25th and 75th percentile (better performance) is statistically significant.

---

**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.***

The Controlling Blood Pressure measure has only one set of specifications.

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

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

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

---

**2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

**2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
HEDIS measures apply to enrolled members in a health plan, and NCQA has a rigorous audit process to ensure the eligible population, denominator, and numerator events for each measure are correctly identified and reported. The audit process is designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.

The HEDIS Compliance Audit addresses the following functions:
- Information practices and control procedures
- Sampling methods and procedures
- Data integrity
- Compliance with HEDIS specifications
- Analytic file production
- Reporting and documentation

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

HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used. Once measures are added to HEDIS, NCQA conducts a first-year analysis to assess the measure's feasibility once widely implemented in the field. This analysis includes an assessment of how many plans report valid rates vs. rates that are materially biased (or have other issues, such as small denominators). These considerations are weighed in the deliberation process before measures are approved for public reporting.

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

The denominator of this measure is identified using claims data and not subject to difference between response or nonresponse. This measure goes through the NCQA audit process each year to identify potential errors or bias in results. Only performances rates that have been reviewed and determined not to be "materially biased" are reported and used.

## 3. Feasibility

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

**3a. Byproduct of Care Processes**

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

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

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

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**.

Some data elements are in defined fields in electronic sources

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

To allow for widespread reporting across health plans and health care practices, this measure is collected through multiple data sources (administrative data, electronic clinical data, and paper records). We anticipate as electronic health records become more widespread, the reliance on paper record review will decrease.

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

**Attachment:**

**3c. Data Collection Strategy**

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

**3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and**

**frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

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

NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO's ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples " comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

1) Information practices and control procedures
2) Sampling methods and procedures
3) Data integrity
4) Compliance with HEDIS specifications
5) Analytic file production
6) Reporting and documentation

In addition to the HEDIS audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system, NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system informs both annual updates to the measures as well as routine re-evaluation of measures. These processes include updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence.

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

Broad public use and dissemination of this measure is encouraged. NCQA has agreed with NQF that noncommercial users do not require the consent of the measure developer. Use by health care providers in connections with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

## 4. Usability and Use

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

**4a. Accountability and Transparency**

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

**4.1. Current and Planned Use**

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

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|
| | Public Reporting<br>Health Plan Ratings<br>https://www.ncqa.org/hedis/reports-and-research/ratings-2019/<br>Health Plan Report Card<br>https://reportcards.ncqa.org/#/health-plans/list<br>Payment Program<br>CMS Medicare Star Rating Program<br>https://www.medicare.gov/find-a-plan/questions/home.aspx<br>CMS Medicaid Adult Core Set<br>https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.html<br>CMS Quality Payment Program:<br>https://qpp.cms.gov/<br>California's Value Based Pay for Performance Program<br>http://www.iha.org/our-work/accountability/value-based-p4p<br>Regulatory and Accreditation Programs<br>NCQA Accreditation<br>https://www.ncqa.org/programs/health-plans/health-plan-accreditation-hpa/<br>Quality Improvement (external benchmarking to organizations)<br>Quality Compass<br>http://www.ncqa.org/hedis-quality-measurement/quality-measurement-products/quality-compass<br>Annual State of Health Care Quality<br>https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report/ |

**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

CALIFORNIA VALUE BASED PAY FOR PERFORMANCE PROGRAM: This measure is used in the California P4P program, which is the largest non-governmental physician incentive program in the United States. Founded in 2001, it is managed by the Integrated Healthcare Association (IHA) on behalf of ten health plans representing 9 million insured persons. IHA reports results on approximately 35,000 physicians in 200 physician organizations.
CMS MEDICARE ADVANTAGE STAR RATING PROGRAM: This measure is included in the composite Medicare Advantage Star Rating. CMS calculates a Star Rating (1-5) for all Medicare Advantage health plans based on 53 performance measures. Medicare beneficiaries can view the star rating and individual measure scores on the CMS Plan Compare website. The Star Rating is also used to calculate bonus payments to health plans with excellent performance. The Medicare Advantage Plan Rating program covers 11.5 million Medicare beneficiaries in 455 health plans across all 50 states.
CMS MEDICAID ADULT CORE SET: There are a core set of health quality measures for Medicaid-enrolled adults. The Medicaid Adult Core Set was identified by the Centers of Medicare & Medicaid (CMS) in partnership with the Agency for Healthcare Research and Quality (AHRQ). The data collected from these measures will help CMS to better understand the quality of health care that adults enrolled in Medicaid receive nationally. Beginning in January 2014 and every three years thereafter, the Secretary is required to report to Congress on the quality of

care received by adults enrolled in Medicaid. Additionally, beginning in September 2014, state data on the adult quality measures will become part of the Secretary's annual report on the quality of care for adults enrolled in Medicaid.
CMS QUALITY PAYMENT PROGRAM: This measure is used in the Quality Payment Program (QPP) which is a reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs).
HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Heath Plans. As of Fall 2017, a total of 184 Medicare Advantage health plans were accredited using this measure among others covering 9.2 million Medicare beneficiaries; 451 commercial health plans covering 113 million lives; and 125 Medicaid health plans covering 35 million lives. Health plans are scored based on performance compared to benchmarks.
HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan rankings which are reported on the NCQA website. These rankings are based on performance on HEDIS measures among other factors. In 2019, a total of 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2018, the report included results from calendar year 2017 for health plans covering a record 136 million people, or 43 percent of the U.S. population.

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

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

N/A

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

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

Health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA. NCQA publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.

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

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Update and Best Practices Conference (now the Health Care Quality Congress), NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly and insight into new measure development projects.

NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c.1.

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

**Describe how feedback was obtained.**

NCQA measures are evaluated regularly using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

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

Questions received through the Policy Clarification Support System have generally centered around clarification on the use of blood pressure readings obtained during potentially stressful procedures or specific visit types, suggestions for exclusions, and the use of patient-reported readings.

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

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

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

During the measure's last major update, feedback obtained through the mechanisms described in 4a2.2.1 informed how we revised the measure, including re-working the denominator approach to minimize burden, adding an administrative approach for the numerator, allowing readings from remote monitoring devices, and updating the numerator threshold to focus on a target of <140/90 mm Hg.

**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.**

Performance across all plan types has generally improved over the past several years, with Medicare, Medicaid, and commercial plan performance increasing each year by about 1%. This trend broke in measurement year 2018 when the measure underwent a number of changes, including changing the denominator to use a purely administrative method for identification of patients with hypertension, the addition of an administrative approach for the numerator, allowing readings from remote monitoring devices to count for the numerator, and changing the numerator threshold to focus on a target of <140/90 mm Hg. As plans adjust to these changes, we expect performance will continue to improve. Current average performance (MY 2018) is highest in Medicare plans (69%), followed by Medicaid plans (59%), and then commercial plans (55%).

**4b2. Unintended Consequences**

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

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

There were no identified unexpected findings during testing or since implementation of this measure.

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

There were no identified unexpected findings during testing or since implementation of this measure.

## 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)**

0061 : Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

2602 : Controlling High Blood Pressure for People with Serious Mental Illness

2606 : Diabetes Care for People with Serious Mental Illness: Blood Pressure Control (<140/90 mm Hg)

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

#0729: Optimal Diabetes Care (NQF Endorsed) - this was not listed in 5.1a.

#0076: Optimal Vascular Care (NQF Endorsed) - this was not listed in 5.1a.

**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.**

There are several related measures that assess blood pressure control but are either focused on different population, use different data sources or are specified at different levels of accountability than NQF 0018. Measure 0061 is NQF endorsed as a single measure that uses health plan reported data to assess the percentage of patients 18-75 years of age with diabetes (type 1 and type 2) whose most recent blood pressure level is <140/90 mm Hg. Measure 2602 is NQF endorsed as a single measure that uses health plan reported data to assess the percentage of patients 18-85 years of age with serious mental illness who had a diagnosis of hypertension and whose blood pressure was adequately controlled during the measurement year. Measure 2606 is NQF endorsed as a single measure that uses health plan reported data to assess the percentage of

patients 18-75 years of age with a serious mental illness and diabetes (type 1 and type 2) whose most recent blood pressure reading during the measurement year is <140/90 mm Hg. Measure 0076 is NQF endorsed as a composite measure (all or nothing) that uses physician reported data to assess the percentage of adult ischemic vascular disease patients, 18-75 years of age, who have optimally managed modifiable risk factors including blood pressure and three other indicators. Measure 0729 is NQF endorsed as a composite measure (all or nothing) that uses physician reported data to assess the percentage of adult diabetes patients, 18-75 years of age, who have optimally managed modifiable risk factors including blood pressure and four other indicators. HARMONIZED MEASURE ELEMENTS: All measures described above focus on a blood pressure target of <140/90 mm Hg. UNHARMONIZED MEASURE ELEMENTS: - Data Source and Level of Accountability: Measures 0018, 0061, 2602, and 2606 are collected through administrative claims and/or medical record review using health plan reported data. Measures 0076 and 0729 are collected through medical record abstraction and reported at the physician level of accountability. - Population Focus: Measure 0018 is focused on the general population of people with hypertension while the other measures focus on either diabetes, serious mental illness with diabetes, or serious mental illness with hypertension. - Age Range: Measures 0018 and 2602 focus on adults 18-85 while the other measures focus on adults 18-75. IMPACT ON INTERPRETABILITY?AND DATA COLLECTION BURDEN:? The differences between measures 0018, 0061, 2602, and 2606 do not have an impact on interpretability of?publicly?reported rates or an impact on data collection burden as the measures are focused on different populations. The differences between 0018, 0076, and 0729 also do not have an impact on interpretability of publicly reported rates or an impact on data collection burden because the data for each measure is collected from different data sources by different entities.

**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.)**

NA

## Appendix

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

Attachment:

## Contact Information

**Co.1 Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance

**Co.2 Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-

**Co.3 Measure Developer if different from Measure Steward:** National Committee for Quality Assurance

**Co.4 Point of Contact:** Brittany, Wade, Wade@ncqa.org

## 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.**

NCQA follows a standard process of vetting members of the measurement advisory panel for conflicts of interest.

CARDIOVASCULAR MEASUREMENT ADVISORY PANEL

Kathy Berra, RN, MSN, ANP-BC, FAHA, FAAN, FPCNA, The LifeCare Company

Donald Casey, MD, MPH, MBA, FACP, FAHA, FAAPL, DFACMQ, American College of Medical Quality

Tom Kottke, MD, MSPH, HealthPartners

Eduardo Ortiz, MD, MPH, Tennessee Valley Healthcare System

Stephen Persell (Chair), MD, MPH, Northwestern University

Randall Stafford, MD, PhD, Stanford University

Kim Williams, MD, MACC, MASNC, FAHA, FESC, Rush University Medical Center

Tracy Wolff, MD, Agency for Healthcare Research and Quality

TECHNICAL MEASUREMENT ADVISORY PANEL

Andy Amster, MSPH, Kaiser Permanente

Sarah Bezeredi, MBA, MSHL, UnitedHealth Group

Jennifer Brudnicki, MBA, Inovalon Inc.

Lindsay Cogan, MS, PhD, New York State Department of Health

Mike Farina, MBA, R.Ph, Capital District Physicians' Health Plan

Marissa Finn, MBA, CIGNA

Scott Fox, MS, Med, FAMIA, The MITRE Corporation

Carlos Hernandez, CenCal Health

Harmon Jordan, ScD, Westat

Gigi Raney, LCSW, Center for Medicaid and CHIP Services

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

Laurie Spoll, Aetna

COMMITTEE ON PERFORMANCE MEASUREMENT

Andrew Baskin, MD, Aetna

Elizabeth Drye, MD, SM, Yale School of Medicine

Andrea Gelzer, MD, MS, FACP, AmeriHealth Caritas

Kate Goodrich, MD, MHS, Centers for Medicare & Medicaid Services

David Grossman, MD, MPH, Washington Permanente Medical Group

Christine Hunter, (Co-Chair), MD, WPS Health Solutions

David Kelley, MD, MPA, Pennsylvania Department of Human Services

Jeffrey Kelman, MMSc, MD, Department of Health and Human Services

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, Freelance

Adrienne Mims, MD, MPH, AGSF, FAAFP, Alliant Health Solutions

Amanda Parsons, MD, MBA, Metroplus

Wayne Rawlins, MD, MBA, ConnectiCare

Misty Roberts, MSN, RN, CPHQ, PMP, Humana

Rudy Saenz, MD, MMM, FACOG, Riverside Medical Clinic

Marcus Thygeson, (Co-Chair), MD, MPH, Blind On-Demand

JoAnn Volk, MA, Georgetown University

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 1999

**Ad.3 Month and Year of most recent revision:** 07, 2018

**Ad.4 What is your frequency for review/update of this measure?** Approximately every 3 years, sooner if the clinical guidelines have changed significantly.

**Ad.5 When is the next scheduled review/update for this measure?** 12, 2020

**Ad.6 Copyright statement:** The HEDIS® measures and specifications were developed by and are owned by the National Committee for Quality Assurance (NCQA). The HEDIS measures and specifications are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties, or endorsement about the quality of any organization or physician that uses or reports performance measures and NCQA has no liability to anyone who relies on such measures or specifications. NCQA holds a copyright in these materials and can rescind or alter these materials at any time. These materials may not be modified by anyone other than NCQA. Anyone desiring to use or reproduce the materials without modification for a non-commercial purpose may do so without obtaining any approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA.



Calculated measure results, based on unadjusted HEDIS specifications, may not be termed “Health Plan HEDIS rates” until they are audited and designated reportable by an NCQA-Certified Auditor. Such unaudited results should be referred to as “Unaudited Health Plan HEDIS Rates.” Accordingly, “Heath Plan HEDIS rate” refers to and assumes a result from an unadjusted HEDIS specification that has been audited by an NCQA-Certified HEDIS Auditor.

Limited proprietary coding is contained in the measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. NCQA disclaims all liability for use or accuracy of any coding contained in the specifications.

Content reproduced with permission from HEDIS, Volume 2: Technical Specifications for Health Plans. To purchase copies of this publication, including the full measures and specifications, contact NCQA Customer Support at 888-275-7585 or visit

www.ncqa.org/publications.

**Ad.7 Disclaimers:** This HEDIS® performance measure is not a clinical guideline and does not establish a standard of medical care and has not been tested for all potential applications.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

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

incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

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 #:** 0071

**Corresponding Measures:**

**De.2. Measure Title:** Persistence of Beta-Blocker Treatment After a Heart Attack

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

**De.3. Brief Description of Measure:** The percentage of patient's 18 years of age and older during the measurement year who were hospitalized and discharged from July 1 of the year prior to the measurement year to June 30 of the measurement year with a diagnosis of acute myocardial infarction (AMI) and who received persistent beta-blocker treatment for six months after discharge.

**1b.1. Developer Rationale:** This measure addresses the appropriate clinical management of a person who has experienced an AMI. Persistent beta-blocker treatment after a heart attack reduces the risk of mortality, reduces the risk of severity of reinfarction, and improves the preservation of the left ventricular function.

**S.4. Numerator Statement:** Patients who received at least 135 days of treatment with beta-blockers during the 180-day measurement interval.

**S.6. Denominator Statement:** An acute inpatient discharge from July 1 of the year prior to the measurement year through June 30 of the measurement year with any diagnosis of acute myocardial infarction (AMI) on the discharge claim.

**S.8. Denominator Exclusions:** Any of the following any time during the patient's history through the end of the continuous enrollment period meet criteria:

- Asthma
- COPD
- Obstructive chronic bronchitis
- Chronic respiratory conditions due to fumes and vapors
- Hypotension, heart block >1 degree or sinus bradycardia
- A medication dispensing event indicative of a history of asthma
- Intolerance or allergy to beta-blocker therapy

Additionally, this measure excludes adults in hospice. It also excludes adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings.

**De.1. Measure Type:** Outcome: Intermediate Clinical Outcome

**S.17. Data Source:** Claims

**S.20. Level of Analysis:** Health Plan

**IF Endorsement Maintenance – Original Endorsement Date:** Aug 10, 2009 **Most Recent Endorsement Date:** Feb 19, 2016

## Preliminary Analysis: Maintenance of Endorsement

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

### Criteria 1: Importance to Measure and Report

#### 1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

**Evidence Summary**

- This is a health plan/integrated delivery system intermediate clinical outcome measure that calculates the percentage of patients 18 years of age and older during the measurement year who were hospitalized and discharged from July 1 of the year prior to the measurement year to June 30 of the measurement year with a diagnosis of acute myocardial infarction (AMI) and who received persistent beta-blocker treatment for six months after discharge. The developer does not describe the evidence for determining the 75% threshold of 180.
- Developer provides decision logic from secondary prevention to intermediate clinical outcome for the persistent use of beta-blockers in reducing the risk of mortality, risk and severity of re-infarction and improving the preservation of the left ventricular function with patients with AMI.
- The developer provides two clinical practice guidelines with four statements for the persistent use of beta-blockers in patients diagnosed with AMI. Grading is provided for each guideline statement including:
    - Beta-blockades during and after hospitalization for a STEMI (Class I/Level B),
    - Beta-blockades in HF patients with reduced systolic function after Non-STEMI/ACS hospitalization (Class I/Level C),
    - Beta-blockades for patients with normal LVF with Non-STEMI/ACS (Class IIa/Level C).

- They also provided a 1999 seminal systematic review summarizing the quantity, quality, and consistency of the evidence, and supporting the recommendations with findings, with Levels B and C.

**Changes to evidence from last review**

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

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

**Updates:** N/A

**Exception to evidence**

N/A

***Questions for the Committee:***

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

**Guidance from the Evidence Algorithm**

Intermediate clinical outcome measure based on systematic review (Box 3) → QQC presented (Box 4) →Quantity: High; Quality: Mod; Consistency: High (Box 5b) →Moderate

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

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

**Maintenance measures – increased emphasis on gap and variation**

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

- The developer provides data from Commercial health plans, Medicare, and Medicaid in the tables below. No additional information is provided on the number or characteristics of the patients included in these data.

**Commercial**

| YEAR | N | MEAN | ST DEV | MIN | 10th | 25th | 50th | 75th | 90th | MAX | IQR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2015 | 245 | 83% | 6% | 62% | 76% | 79% | 83% | 88% | 91% | 99% | 9% |
| 2016 | 251 | 84% | 7% | 57% | 76% | 80% | 85% | 89% | 92% | 98% | 9% |
| 2017 | 243 | 85% | 6% | 57% | 77% | 81% | 85% | 89% | 92% | 100% | 8% |

**Medicaid**

| YEAR | N | MEAN | ST DEV | MIN | 10th | 25th | 50th | 75th | 90th | MAX | IQR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2015 | 115 | 80% | 11% | 43% | 64% | 75% | 83% | 88% | 92% | 97% | 13% |
| 2016 | 136 | 80% | 9% | 50% | 67% | 77% | 81% | 86% | 90% | 95% | 9% |
| 2017 | 145 | 78% | 9% | 39% | 66% | 74% | 80% | 84% | 89% | 97% | 10% |

**Medicare**

| YEAR | N | MEAN | ST DEV | MIN | 10th | 25th | 50th | 75th | 90th | MAX | IQR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2015 | 258 | 91% | 5% | 68% | 85% | 88% | 91% | 94% | 97% | 100% | 6% |
| 2016 | 256 | 90% | 5% | 61% | 83% | 88% | 91% | 94% | 96% | 100% | 6% |
| 2017 | 272 | 90% | 5% | 71% | 84% | 88% | 91% | 93% | 95% | 100% | 6% |

- Rates for Commercial, Medicaid, and Medicare appear relatively unchanged annually. Further explanation of the performance data characteristics and findings is not provided.

**Disparities**

- The developer provides a summary of the 2019 CMS Racial, Ethnic, and Gender Disparities in Health Care in Medicare Advantage report describing the racial and ethnic disparities among beneficiaries 18-years and older who received persistent beta-blocker treatment for six months following a hospital discharge for a heart attack. The report infers race and ethnicity and reports the following results:
    - White: 92.2% received treatment
    - Asian or Pacific Islander: 90% received treatment
    - Black: 86.8% received treatment
    - Hispanic: 87.6% received treatment
- The developer states they do not collect performance data stratified by race, ethnicity, or language, though other HEDIS measure sets are available for stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities.
- The developer summarizes data from the literature on the prevalence of heart disease, medication adherence among MI survivors by disability, status, race/ethnicity, and income for all Medicare fee-for-service beneficiaries and the impact of employment status on rates of CHD/stroke. The summary demonstrates disparities in premature death due to heart disease or stroke and in rates of recurrent MI or fatal CHD.

***Questions for the Committee:***

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

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

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

1a.

- Intermediate outcome; evidence based on SR, QQC and evidence graded
- The data has significantly changed since this measure was proposed and endorsed. There is now data that benefit of beta blocker therapy in certain types of AMI receiving contemporary therapies may be limited or non-existent. There are newer studies that raise real questions.
- I am aligned the research hasn't drastically changed; however, it seems like the steward is calculating adherence, rather than persistence. As noted the 75% adherence threshold is not mentioned either.
- Evidence clear and unchanged
- No changes to the evidence

1b.

- There is a performance gap of ~15%; disparities data are available and show a gap of about 5 %age points for race and Medicaid in 2017
- There is a gap, but 100% is not a reasonable clinical goal due to inclusion criteria being too broad not tightly enough defined
- Some disparity data noted. Gap exists for measure. Concerned there is no change over the three years examined.
- Present within plans and across racial groups
- 50th percnetile is about 80-85%, there is room for improvement.

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability: Specifications and Testing**

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

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

### Reliability

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

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

### Validity

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

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

**Composite measures only:**

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

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

**Evaluators:**

Methods Panel Review (Combined)

**Methods Panel Evaluation Summary**:

This measure was reviewed by the Scientific Methods Panel Subgroup. A summary of the measure and the Subgroup discussion is provided below. The full Scientific Methods Panel did not pull this measure for discussion, accepting the ratings of the Subgroup.

Reliability: H-2; M-5; L-0; I-0

- Testing data included HEDIS 2018 plan data (2017 measurement year)
    - 243 commercial plans, median denominator size=65
    - 145 Medicaid plans, median denominator size=81
    - 272 Medicare plans, median denominator size=72
- Score-level reliability testing was conducted using the beta-binomial model described by Adams (2009).
    - Average reliability, commercial: 0.757; 25th percentile=0.521, median=0.672
    - Average reliability, Medicaid: 0.818; 25th percentile=0.389, median=0.621
    - Average reliability, Medicare: 0.730; 25th percentile=0.670, median=0.772
- Data element reliability testing was not conducted. NOTE that such testing is NOT required by NQF for this type of measure.

Validity: H-0; M-5; L-1; I-1

- Testing data same as described above
- Score-level construct validation was conducted by correlating the scores for this measure to those of a measure of statin therapy adherence.
    - Developers hypothesized that a plan that does well on the statin adherence measure for cardiovascular disease would also do well on this measure.
        - Pearson correlation coefficient, commercial: 0.51 (statistically significant)
        - Pearson correlation coefficient, Medicaid: 0.60 (statistically significant)
        - Pearson correlation coefficient, Medicare: 0.42 (statistically significant)
        - Developers interpret these results as supporting their hypothesis and validating this measure.
- Exclusions
    - This measure includes several exclusions, but except for advanced illness/frailty, the developers did not test the exclusions.
    - Exclusions related to hospice enrollment, I-SNP enrollment, living in a long-term care intuitional setting, and advanced illness and frailty are new since the measure was last evaluated by NQF for endorsement.
        - Exclusions for advanced illness resulted in a loss of 4.6% of patients on average (in 10 plans), and a 2.5% higher performance rate on average across the 10 plans
        - Exclusions for frailty resulted in a loss of 1.1% of patients on average (in 10 plans), and a 0.5% higher performance rate on average across the 10 plans
- This measure is not risk-adjusted. Developers provide a brief conceptual rationale regarding lack of risk-adjustment.
- To demonstrate ability to identify statistically meaningful differences across health plans:
    - The developers presented distributional statistics by plan type (e.g., average, standard deviations, IQR, etc.)
    - Used an independent sample t-test of the performance difference between two randomly selected plans at the 25th versus 75th percentile. The test

statistic is compared against a normal distribution. If the p-value of the test statistic is less than .05, then the two plans' performance is significantly different from each other. P-values for all three plan types were <0.05.

- Missing data
  - The developers describe how their audit process considers missing data. However, they do not present information on the extent of missing data.

- Some concerns of the SMP
  - Unclear why the developers have classified this measure as an intermediate clinical outcome
  - Desire for clarity regarding patients who have <180 days of treatment because they died in that timeframe
  - Would like to have seen more detail regarding the method used to test reliability, as well as more detail regarding reliability in relation to sample size
  - Concern regarding low reliability for many plans (particularly Medicaid plans)
  - Desire for data regarding frequency of exclusions
  - Desire for data characterizing the extent of missing data
  - There is some disagreement about the need for risk-adjustment, given the characterization of the measure as an intermediate clinical outcome (note that risk-adjustment not expected for process measures)
  - Some concern about the utility of the statin measure as a comparator/validator for this measure

***Questions for the Committee regarding reliability:***

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

***Questions for the Committee regarding validity:***

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

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

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

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

2a1.

- The measure was rated and passed by the sci methods panel
- The diagnosis of AMI has greatly evolved in the modern day due to widespread use of troponin and high sensitivity troponins, altering the denominator. Also, there are significant coding pressures to code AMI in borderline clinical situations. Unclear if Type II MIs are included. If measure denominaotr is kept as proposed, risk adjustment for EF and CHF is definitely needed.
- Reviewed by Scientific Methods Panel. Subcommittee approved
- Moderate
- Score-level testing with adequate reliabiliting using signal to noise. Scores are reasonable (0.62-0.77

2a2.

- No
- This is a process measure, not an intermediate clinical outcome
- Reviewed by subcommittee.
- Moderate
- no

2b1.

- Modest concern about statins as a comparator
- see reliability comment above (6.2a)
- Pearson scores seemed low for the 0.7 threshold
- So-so correlation with statin measure results for Medicare plans in general; Medicare beneficiaries would be most likely to have AMI of any population
- Score-level construct validation with correlation coefficients ranging from 0.42-0.60

2b4-7.

- I don't see any significant threats to validity
- See above comments re evidence etc.
- No concerns.
- No major threats noted
- testing with and without the frailty exclusion showed a slight change in performance

2b2-3.

- Exclusions are appropriate
- Exclusions of COPD and Obstructive chronic bronchitis are not evidence based.
- Denominator exclusion includes individuals dispensed a medication for the treatment of dementia. Unsure why this is an exclusion
- I do not believe there is adjustment for race
- Not risk-adjusted

## Criterion 3. Feasibility

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

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

- The developer states all data elements are in defined fields in electronic claims.
- The developer states the data elements are generated, collected, and used by healthcare personnel during the provision of care and are coded and abstracted by someone other than the person obtaining original information.
- The developers do not provide specific information on the operational use of the measure; instead they outline the HEDIS Compliance Audit process to verify that HEDIS specifications are met. In addition to the audit, NCQA provides a system that allows for 'real-time' feedback from measure users.

***Questions for the Committee:***

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?

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

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

- I agree that feasibility is moderate
- Claims based, patients may have received beta blockers without an Rx claim. Some clinical record review would be appropriate.
- No concerns
- no concerns
- Seems feasible

## Criterion 4: Usability and Use

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

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

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

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

**Current uses of the measure**

**Publicly reported?** ☒ **Yes** ☐ **No**

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

**OR**

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

**Accountability program details**

- The developer states that this measure is publicly reported in NCQA's State of Health Care annual report and Quality Compass. It is also used to calculate health plan rankings reported in Consumer Reports. This measure is also used in scoring for accreditation of Medicare Advantage Health Plans.

**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 states feedback received through the Policy Clarification Support System include:

- Clarification questions on specific language used
- Suggestions for potential exclusions
- Clarifications on recently added exclusion for advanced illness and frailty

**Additional Feedback:** N/A

***Questions for the Committee**:*

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

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

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

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

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

**Improvement results**

- The developer states over the past three years:
  - Commercial plan performance has increased annually by 1%
  - Medicare plan performance has remained relatively stable
  - Medicaid plan performance decreased by 2%
- Per the developer, current average performance is highest in Medicare plans, followed by commercial plans, and then Medicaid plans. Developer states there is still room for improvement despite encouraging performance – no additional data was provided.

**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**

- Developer states there were no unexpected findings.

**Potential harms**

- Developer does not provide potential harms.

**Additional Feedback:** N/A

***Questions for the Committee**:*

- Are you aware of any unintended consequences of this measure?
- Do the benefits of the measure outweigh any potential unintended consequences?

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

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

4a1.

- This measure is publicly reported in NCQA's State of Health Care annual report and Quality Compass. Also used in health plan rankings.
- It is used.
- Used within Public reporting, health plan ratings, and NCQA accredition.
- Unclear if currently used in accountability program
- Currently publicly reported

4b1.

- No significant harms identified. I agree with moderate
- Beta blockers are less clinically importnat in polst MI treatemnt than anto paltelets and statins, this measure may promote a less effective treatment over a more effective one(s). May promote a clinically ineffective treatment in a large subset.
- No issues
- No concerns
- No concerns

## Criterion 5: Related and Competing Measures

**Related or competing measures**

- 0070: Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF < 40%)

**Harmonization**

- Developer states measure specifications are not harmonized to the extent possible as this measure focuses on beta-blocker treatment post-AMI, while measure 0070 focuses on patients who have a prior MI or a current or prior LVEF < 40%. Furthermore, data is collected from different sources for the two measures and the LOA of the two measures differs. Measure 0070 also has different exclusion criteria.

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

- 0070 is related but not competing.
- Measure 0070 is a more appropriate measure in general
- It may be beneficial to better align the exclusions between the two measures.
- No concerns
- related measure incorporates measurement of LVEF <40%

## Public and Member Comments

**Comments and Member Support/Non-Support Submitted as of: Month/Day/Year**

- **Of the XXX NQF members who have submitted a support/non-support choice:**
  - XX support the measure

- YY do not support the measure

## Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

**Measure Number:** 0071

**Measure Title:** Persistence of Beta-Blocker Treatment After a Heart Attack

**Type of measure:**

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**
☐ **Outcome** ☐ **Outcome: PRO-PM** ☒ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

**Panel Member #4:** *The provided materials refer to the measure as an "Outcome: Intermediate Clinical Outcome," however, NQF materials specify that such a measure is "a change in a physiologic state that leads to a longer-term health outcome." I do not see any description of the measure satisfying this criterion. Moreover, this measure was previously described as a process measure in NQF documents. Perhaps to frame this as an outcome, the argument could be made that medication adherence is a "heath-related behavior." This argument is not made. I am evaluating this as a process measure.*

**Panel Member #6:** *NOTE: The developer says that this is an "Outcome: Intermediate Clinical Outcome" measure, but I believe that it is a process measure since it has to do with the receipt of medication.*

**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.)

**Panel Member #4:** *Previously endorsed as a process measure, NOT an outcome.*

### RELIABILITY: SPECIFICATIONS

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

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

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

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

**Panel Member #1:** Specifications are clearly defined. I have no concerns.

Not being familiar enough with the specifics of the claims data used, I was wondering if there is a way that NCQA could also track actual medication purchases in addition to prescriptions, to better address actual patient compliance, acknowledging a purchase is still not evidence for full compliance, but more so than evidence on prescriptions. If possible, this would allow a better measurement of the continuity of treatment, rather than the continuity of prescribed treatment.

**Panel Member #2:** None. There are some out-of-date dates (Ad.2-Ad.5) which raises the concern of general accuracy of the document.

**Panel Member #4:** *(Note above re: description as outcome measure rather than process measure).*

*I cannot claim expertise on measurement of outpatient medication adherence, so I defer. Although unlikely to have a large effect, the 180-day interval includes the data of discharge (plus 179 days afterwards). Some patients may receive oral beta blocker therapy as inpatients on the day of discharge, thus undercounting days of outpatient therapy by 1 day.*

**Panel Member #5:** The specifications are clear and detailed.

**Panel Member #6:** *None*

**Panel Member #7:** The measure specifications lack one very important detail. The specifications do not indicate how the measure is to be estimated if the patient has less than 180 days of follow-up after discharge—notably, due to death. Death after MI discharge within months after AMI discharge is not uncommon.

### RELIABILITY: TESTING

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

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

   **Panel Member #1:** The methods used for score reliability testing cannot be fully assessed using the information provided. Only the general concepts of the signal to noise ratio were described, referencing the Beta-binomial model (Adams, 2009), but no details were provided on the actual methods and formulas used.

   More details on the statistical method and specific formulas used to calculate the proportion of variability in measured performance that can be explained by real differences in performance would be helpful to better understand what was done exactly.

   **Panel Member #2:** The developer uses a common approach the Beta-binomial model (Adams 2009) to estimate signal-to-noise and reports the reliability statistic distribution (although not stratified by volume which would be more informative since median denominator size is relatively small (65-81)).

   **Panel Member #3:** Tested SNR using the beta-binomial test.

   **Panel Member #4:** *Beta-binomial calculation (within product line, across about 100 or more plans within each, but with median denominator size of 65-81 patients) is presumably adequate.*

**Panel Member #5:** Score-level reliability was assessed with a Beta-binomial model (Adams, 2009), stratified by insurance type. This is a standard signal-to-noise analyses.

**Panel Member #6***: For Measure score level testing, the beta-binomial method is appropriate.*

**Panel Member #7:** The testing is appropriate. The steward fit a beta-binomial regression model to the plan-level measure values, thus permitting estimation of beta distribution parameters. From this, reliability estimates can be readily derived, according to well-known statistical methodology.

**Submission document:** Testing attachment, section 2a2.2

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

**Panel Member #1:** These results seem to be related to the sample size used per product line. For example, Medicaid had the highest median number of cases and highest overall reliability. I recommend also reporting the minimum number of patients per health plan needed to achieve acceptable levels of overall reliability. Health plans that do not reach that minimum threshold might be difficult to assess due to their low score level reliability.

**Panel Member #2:** The developer reports that the reliability for half of the health plans across product lines demonstrate high reliability. However half demonstrate moderate reliability, with 10% demonstrating low reliability, especially in Medicaid health plans.

**Panel Member #3:** SNR ranged between 0.74 and 0.82. Half of the health plans had SNR >= 0.7. This is acceptable.

**Panel Member #4:** *Reliability estimates are 0.739 to 0.818. The distributions are reported such that "approximately half of health plans (across all product lines) are either right at the threshold of 0.7 or exceed it."*

**Panel Member #5:** The overall reliability was good for all three insurance types (>0.70). At the plan-level, roughly 50% of the plans have reliabilities <0.70 with some very low values especially for medicare and medicaid plans. A description of the relationship between plan reliability and denominator size might help in understanding the influence of low sample size.

**Panel Member #6:** *The results strongly demonstrate reliability.*

**Panel Member #7:** Overall reliability is good in commercial, Medicaid, and Medicare health plans, but it is also clear that reliability declines greatly when the number of qualifying AMI discharges in a plan is low. In commercial and Medicaid plans, where average age can be expected to be lower (relative to Medicare plans), and absolute risk of AMI is thus lower, the first quartile of reliability is quite low.

**Submission document:** Testing attachment, section 2a2.3

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

**Submission document:** Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

**Submission document:** Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

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

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

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

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

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

**Panel Member #1:** I would have preferred that more details be provided on actual formulas used to calculate the between & within variance for the signal to noise testing. As noted above, identifying the minimum number of patients per health plan needed to achieve acceptable levels of overall reliability would be informative as health plans with too few cases might not be reliably assessed.

The high rating is due to my understanding that these details are currently not an NQF requirement.

**Panel Member #2:** For the level of reporting (i.e. health plan by product line with varying sample size) the reliability must be rated as moderate. Reporting reliability results using an alternative methodology (e.g. ICC) might increase confidence in the application of Beta-binomial model since several of the health plans demonstrate perfect reliability (a metric of 1.0) which seems unlikely with a denominator less than 100.

**Panel Member #3:** SNR ranged between 0.74 and 0.82. Half of the health plans had SNR >= 0.7. This is acceptable

**Panel Member #4:** *Overall reliability varies within product line, and one-half of health plans are below 0.7. I seek others' opinions, however, my impression at this time is that such performance would fit low-moderate reliability at best.*

**Panel Member #5:** Methodology for reliability testing was strong. The results were good overall but raised some questions about plans with low reliability, probably due to low sample sizes. Some more descriptive data on the distribution of sample sizes (not just the median) and the relationship between size and reliability would be informative.

**Panel Member #6:** *Appropriate methods strongly demonstrate reliability.*

**Panel Member #7:** The measure is generally reliable, but nevertheless prone to low reliability in non-Medicare health plans.

## VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

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

**Panel Member #1:** Clinical justifications for the exclusions selected were supported by expert panels but were mostly un-tested, except for testing of a sub-sample for the exclusion of patients with advanced illness and frailty. I recommend that at least information on frequency of all excluded populations be provided. Depending on the extent to which important sections of the populations are excluded, using a risk-adjustment approach rather than excluding patients could be considered given the pros and cons of each option.

A similar discussion on potential benefits or risks of excluding vs. risk-adjusting for patient with advanced illness and frailty would be informative. For example, 5% of patients aged 66-80 were excluded due to advanced illness and frailty, with some, but minor impact on performance rates.

Adjusting for advanced illness and frailty might offer an opportunity to include these patients in this measure. Was this option considered? If yes, what were the considerations for excluding this patient group and not adjusting for their illness status?

**Panel Member #2:** In general the stated exclusions are well-rationalized and the lengthy discussion of advanced illness and frailty seems well justified and the percent excluded (~1-4%) small. As always more data are better and reporting of the rate stratified by with and without the exclusion was informative.

**Panel Member #3:** none

**Panel Member #4:** *The provided materials state, "we did not perform testing of … exclusions…"*

**Panel Member #5:** None

**Panel Member #6:** *None*

**Panel Member #7:** I have no concerns.

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

**Submission document:** Testing attachment, section 2b4.

**Panel Member #1:** Inter-quartile ranges were between 5-10%, with lowest rates being relatively high, especially for the Medicare population (71%). This raises questions about the potential for a topped-out measure and challenges the developer's interpretation of the statistical differences translating into meaningful differences in performance. I think this supports the recommendation to move toward a more valid measure of actual treatment adherence rather than assessing only evidence on prescriptions.

**Panel Member #2:** Given how long this measure has been in use the more relevant question might be the ability to identify meaningful differences in performance over time. Is there any evidence that health plans are able to improve the quality of care? The IQR although significant are relatively modest in magnitude, lessening the utility of the measure.

**Panel Member #3:** none

**Panel Member #4:** *No significant concerns – ability to identify differences between 25$^{th}$ and 75$^{th}$ percentiles, and such differences could be considered meaningful (6-8% absolute differences).*

**Panel Member #5:** Especially for Medicare, the 10%ile of the performance distribution is 83.7% with an IQR of only 5.7%. The t-tests comparing the 25$^{th}$ and 75$^{th}$ percentiles is less informative than how clinically meaningful it would be for a plan to move from 87.7% to 93.4% (for medicare). This is a question for the Standing Committee. The distributions in the other insurance lines are less compressed but still raise questions about ceiling effects.

**Panel Member #6:** *No concerns.*

**Panel Member #7:** I have no concerns.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

**Submission document:** Testing attachment, section 2b5.

**Panel Member #1:** Only one set of specifications was used.
However, in section 1.6, no descriptive characteristics of the patients included are provided. Additionally, providing the range (minimum & Maximum) number of patients per plan would be informative.

Also, in section 1.7, exclusions were tested on a sub sample that was not described or compared to the overall sample.

**Panel Member #2:** Not applicable (although one might wonder whether the administrative data and/or medical chart review is relevant).

**Panel Member #4:** *No significant concerns.*

**Panel Member #5:** None

**Panel Member #6:** *No concerns.*

Panel Member #7: This item is not applicable.

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

**Submission document:** Testing attachment, section 2b6.

**Panel Member #1:** There is a general description about audits that "verify primary data sources used to populate measures and ensure specifications are correctly implemented."
"If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used."
No specific results are provided about rates of missing data and number of health plans excluded due to missing data. I suggest this information be added. If the rate of excluded plans is not negligible, some testing on threats to validity due to these exclusions is warranted.

**Panel Member #2:** The HEDIS data are subject to systematic audit which is one of the advantages of the HEDIS measurement system.

**Panel Member #3:** Testing for missing data not identified.

**Panel Member #4:** *No significant concerns.*

**Panel Member #5:** None. NCQA has a good system to assess and address missing data. Less of a concern with claims data.

**Panel Member #6:** *No concerns.*

**Panel Member #7:** I have no concerns.

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

**Panel Member #4:** *This is subject to debate.*

**Panel Member #6:** *Because the proposed measure is a process rather than an outcome measure, there is no need for a rationale to not risk adjust. Indeed, the rationale given in 2b3.2 makes the case that this really is a process measure.*

16c. **Social risk adjustment:**

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

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

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

16d. **Risk adjustment summary:**

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

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

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

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

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

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

**Panel Member #1:** Risk-adjustment was not part of this measure.

**Panel Member #2:** Given the amount of variation in performance across plan type (esp. Medicaid) and across health plans shown in Table 4 it is difficult to conclude that all of that variation is attributable to quality of care (and also the high estimates of reliability) rather than patient factors or other contextual factors (urban or rural).

**Panel Member #3:** Not applicable

**Panel Member #5:** No risk adjustment and a good rationale for not doing so.

**For cost/resource use measures ONLY:**

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

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

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

**VALIDITY: TESTING**

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

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

**☐ Face validity**

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

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

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

**Panel Member #1:** A correlation between measure 0071 and what was defined as a similar measure, i.e., receiving statin medication, does not necessarily support its validity, as I assume there may be clinical reasons for prescribing beta-blockers but not prescribing statins, or vice versa. Similarly, a low correlation between these measures does not necessarily prove the measure is not valid. However, I appreciate the fact that identifying a more similar comparative measure other than the one selected here might be challenging.

I'd like to suggest a different avenue for validity testing.

Measure 0071 assesses the continuity of beta-blocker treatment after a heart attack. It does so by assessing the continuity of prescriptions filled during the measurement period (180 days), assuming prescriptions are translated into actual treatment. This assumption is known to be challenging as prescriptions do not necessarily translate into actual treatment. I wonder if the developers could have a way to support that the measure of continuity of beta-blocker prescriptions does in fact measure the continuity of beta-blocker treatment, by correlating scores to a secondary medication adherence measure, e.g., medication acquisition, clinician assessment or patient self-report?

**Panel Member #2:** The developer examines the Pearson correlation among related measures (persistence of BB treatment after heart attack and stain therapy for patients with cardiovascular disease). The stated implicit quality construct is "medication adherence during a specified time frame".

**Panel Member #3:** Assessed construct validity using measure for statin adherence. Pearson correlation coefficient for 3 groups (Medicare, Medicaid, private) ranged between .4 and 0.6. This is acceptable.

**Panel Member #4:** *Construct validity with statin in atherosclerosis measure.*

Panel Member #5: Validity was assessed by calculating the Pearson correlation of plans' performance on this measure with another measure of statin persistence for patients with CVD.

**Panel Member #6:** *The method (correlating the proposed measure with a measure of statin therapy adherence) is appropriate*

**Panel Member #7:** The steward assessed whether the measure was correlated with statin adherence among patients with atherosclerotic cardiovascular disease.

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

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

**Panel Member #1:** As reported, correlations were in the moderate range, with large differences between product lines (0.4-0.6). This could be interpreted as moderate evidence for validity. However, as noted above, moderate correlations might also result from clinical considerations when only one of the two drugs were clinically appropriate, clinical side effects, etc. Thus, additional methods to assess validity should be considered.

**Panel Member #2:** The developer reports correlation among component measures in excess of 0.42-0.60 across plan types at the measured entity level. Correlations of this magnitude are considered moderate.

**Panel Member #3:** See above

**Panel Member #4:** *Moderate correlation (within all health plans, 0.42 to 0.60) between these two measures.*

**Panel Member #5:** I agree that "across all product lines, the correlations are moderate and statistically significant, which suggests plan performance on the Statin Therapy for Patients with Cardiovascular Disease - Adherence 80% measure is correlated to performance on the Persistence of Beta-Blocker Treatment After a Heart Attack measure. Plans that have higher rates on one measure will have higher rates on the other." A scatterplot of the pairs performance scores and a confidence interval for the correlations would be helpful.

**Panel Member #6:** *The correlations are moderate (rho = 0.51, 0.60 & 0.42), which is what one would expect.*

**Panel Member #7:** Correlations were positive, but modest. This analysis is supportive of the validity of the measure, but does not provide direct evidence of validity.

**Submission document: Testing attachment, section 2b2.3**

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

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

☒ **Yes**

☐ **No**

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

**Panel Member #4:** *I do not have confidence in my response above. I am uncertain whether this is adequate and will appreciate learning more about this and alternative approaches.*

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

*NOTE that data element validation from the literature is acceptable.*

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

☐ **Yes**

☐ **No**

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

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

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

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

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

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

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

**Panel Member #1:** As noted, I believe alternative measure of validity should be considered, e.g., establishing a relationship between patient groups with different levels of expected adherence, or other measures of treatment adherence to beta-blockers. However, I also agree the approach used is reasonable enough to be accepted, thus the moderate rating.

Additionally, there are potential threats to validity that need to be addressed before a 'high' rating can be given, including:

- A more comprehensive reporting of excluded populations and exclusion testing.
- Considerations about risk-adjustment as an alternative to some of the exclusion parameters.
- Concerns about the possibility of a topped-out measure, especially for the Medicare group.
- Additional information on the extent of missing data.

**Panel Member #2:** A demonstration of an implicit quality construct is the lowest level of empirical validity testing. To demonstrate a moderate level, the developer must show an empirical association between the implicit quality construct and the material outcome (or better yet an explicit quality construct and the material outcome). For example, that health plans with worse performance on medication adherence have worse performance on mortality, reinfarction, and left ventricular function.

**Panel Member #3:** Acceptable level of construct validity.

**Panel Member #4:** *The nature of this measure (process or intermediate clinical outcome) must be established.*

*The numerator and denominator statements are established in this measure, however, my own understanding of how they relate to patients taking pills is limited.*

*I will appreciate learning the perspectives of others as to whether the range of reliability estimates or the approach to construct validity constitute "potential threats" to the measure.*

**Panel Member #5:** The distributions of performance are high and a bit compressed. I'm wondering at what point victory can be declared.

The analysis of concurrent validity with a similar measures is good. A stronger test of the measure's validity would be if patients who meet the measures have better outcomes than patients who do not meet the measure (predictive validity). I understand that the sponsor might not have the patient-level data to do those analyses.

**Panel Member #6:** *Correlating the proposed measure with one other measure is appropriate, but not enough to convince me that the validity is high. I would like to see the proposed measure compared in this way to more measures.*

**Panel Member #7:** Numerator compliance in the measure is clear: at least 75% adherence (proportion of days covered) during the 180 days following discharge. However, what remains unclear to me is how the measure classifies patients with less than 180 days of follow-up after discharge.

### FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

### ADDITIONAL RECOMMENDATIONS

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

    **Panel Member #1:** I believe the SMP should discuss these concerns before forwarding them to the standing committee.

    **Panel Member #2:** Note to NQF staff: we need an interpretative standard for reliability metrics and pearson correlations rather than having each developer cite a standard (or at least cite the authority for the standard).

## 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**

PBH_Evidence_Form_-71-.docx

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

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

No

### 1a. Evidence (subcriterion 1a)

**Measure Number** (*if previously endorsed*)**:** 0071

**Measure Title**: Persistence of Beta-Blocker Treatment After a Heart Attack

**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**: 8/1/2019

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

Outcome

☐ Outcome: Click here to name the health outcome

☐ 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*): A 180-day course of treatment with beta blockers

☐ 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.

Patient 18 years of age and older is hospitalized>>> Health care provider diagnoses patient with acute myocardial infarction (AMI)>>> Health care provider and patient discuss the risk and benefits of beta-blocker therapy post discharge>>> Patient is dispensed a 180-day course of treatment with beta-

blockers>>> Persistent beta-blocker use in patient's treatment reduces the risk of mortality, reduces the risk and severity of reinfarction, and improves the preservation of the left ventricular function>>> Improvement in quality of life and functioning for patient (Desired outcome).

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

N/A

****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.**

N/A

**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

Table 1. ST-Elevation Myocardial Infarction (STEMI) Guideline

| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | • **Guideline for the Management of ST-Elevation Myocardial Infarction**<br>• **American College of Cardiology Foundation/American Heart Association**<br>• **January 2013**<br>• **J Am Coll Cardiol. 2013;61(4):e78-e140. doi:10.1016/j.jacc.2012.11.019**<br>• **URL: http://content.onlinejacc.org/article.aspx?articleid=1486115** |

| | |
|---|---|
| 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. | Recommendation:<br>• "Beta blockers should be continued during and after hospitalization for all patients with STEMI and with no contradictions to their use." (Level B; Class I) |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | Level of Evidence & Description:<br>Level B<br>• Limited populations evaluated<br>• Data derived from a single randomized trial or nonrandomized studies<br>• Evidence from single randomized trial or nonrandomized studies<br>• Some conflicting evidence from single randomized trial or nonrandomized studies<br>• Greater conflicting evidence from single randomized trial or nonrandomized studies<br>• Evidence from single randomized trial or nonrandomized studies |
| Provide all other grades and definitions from the evidence grading system | Level of Evidence & Description:<br>Level A<br>• Multiple populations evaluated<br>• Data derived from multiple randomized clinical trials or meta-analyses<br>• Sufficient evidence from multiple randomized trials or meta-analyses<br>• Some conflicting evidence from multiple randomized trials or meta-analyses<br>• Greater conflicting evidence from multiple randomized trials or meta-analyses<br>• Sufficient evidence from multiple randomized trials or meta-analyses<br>Level C<br>• Very limited populations evaluated<br>• Only consensus opinion of experts, case studies, or standard of care<br>• Only expert opinion, case studies, or standard of care<br>• Only diverging expert opinion, case studies, or standard of care<br>• Only expert opinion, case studies, or standard of care |
| Grade assigned to the **recommendation** with definition of the grade | The grades assigned by the ACC/AHA to the guideline varied by the guideline recommendation. See question above for the grade given to each guideline recommendation.<br><br>Class of Recommendation: |

| | |
|---|---|
| | Class I<br>• Benefit > > > Risk<br>• Procedure/Treatment SHOULD be performed/administered<br>• Recommendation that procedure or treatment is useful/effective |
| Provide all other grades and definitions from the recommendation grading system | Class of Recommendation:<br>Class IIa<br>• Benefit > > Risk<br>• Additional studies with focused objective needed<br>• IT IS REASONABLE to perform procedure/administer treatment<br>• Recommendation in favor of treatment or procedure being useful/effective<br>Class IIb<br>• Benefit ≥ Risk<br>• Additional studies with broad objectives needed; additional registry data would be helpful<br>• Procedure/Treatment MAY BE CONSIDERED<br>• Recommendation's usefulness/efficacy less well established<br>Class III – No benefit<br>• Procedure/Test – Not Helpful<br>• Treatment – No proven benefit<br>• Recommendation that procedure or treatment is not useful/effective and may be harmful<br>Class III – Harm<br>• Procedure/Test – Excess cost without benefit or harmful<br>• Treatment – Harmful to patients<br>• Recommendation that procedure or treatment is not useful/effective and may be harmful |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | The ACC/AHA does not provide information on the systematic review conducted to support its guideline and the recommendations mentioned above. In lieu of the ACC/AHA systematic review, we reported on another systematic review of the **evidence** that supports the ACC/AHA recommendations in Table 3. |
| Estimates of benefit and consistency across studies | See Table 3. |
| What harms were identified? | See Table 3. |
| Identify any new studies conducted since the SR. Do the | There have been no new studies that contradict the current body of evidence. |

| | |
|---|---|
| new studies change the conclusions from the SR? | |

Table 2. Non-ST Elevation Myocardial Infarction (NSTEMI) Guideline

| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | • **Guideline for the Management of Patients with Non-ST-Elevation Acute Coronary Syndromes**<br>• **American College of Cardiology/American Heart Association**<br>• **December 2014**<br>• **J Am Coll Cardiol. 2014;64(24):2645-2687. doi:10.1016/j.jacc.2014.09.016.**<br>• **URL: http://content.onlinejacc.org/article.aspx?articleid=1910085&resultClick=3** |
| 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. | Recommendations:<br>• "In patients with concomitant NSTE-ACS [non-ST-elevation acute coronary syndrome], stabilized HF [heart failure], and reduced systolic function, it is recommended to continue beta blocker therapy with 1 of the 3 drugs proven to reduce mortality in patients with HF: sustained-release metoprolol succinate, carvedilol, or bosoprolol." (Level C; Class I)<br>• "It is reasonable to continue beta blocker therapy in patients with normal LV [left ventricular] function with NSTE-ACS" (Level C; Class IIa)<br>• "Medications required in the hospital to control ischemia should be continued after hospital discharge in patients with NSTE-ACS who do not undergo coronary revascularization, patients with incomplete or unsuccessful revascularization, and patients with recurrent symptoms after revascularization. Titration of the doses may be required." (Level C; Class I) |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | Level C<br>• Very limited populations evaluated<br>• Only consensus opinion of experts, case studies, or standard of care<br>• Only expert opinion, case studies, or standard of care<br>• Only diverging expert opinion, case studies, or standard of care<br>• Only expert opinion, case studies, or standard of care |
| Provide all other grades and definitions from the evidence grading system | Level of Evidence & Description:<br>Level A<br>• Multiple populations evaluated<br>• Data derived from multiple randomized clinical trials or meta-analyses<br>• Sufficient evidence from multiple randomized trials or meta-analyses |

| | |
|---|---|
| | • Some conflicting evidence from multiple randomized trials or meta-analyses<br>• Greater conflicting evidence from multiple randomized trials or meta-analyses<br>• Sufficient evidence from multiple randomized trials or meta-analyses<br>Level B<br>• Limited populations evaluated<br>• Data derived from a single randomized trial or nonrandomized studies<br>• Evidence from single randomized trial or nonrandomized studies<br>• Some conflicting evidence from single randomized trial or nonrandomized studies<br>• Greater conflicting evidence from single randomized trial or nonrandomized studies<br>• Evidence from single randomized trial or nonrandomized studies |
| Grade assigned to the **recommendation** with definition of the grade | The grades assigned by the ACC/AHA to the guideline varied by the guideline recommendation. See question above for the grade given to each guideline recommendation.<br><br>Class of Recommendation:<br>Class I<br>• Benefit > > > Risk<br>• Procedure/Treatment SHOULD be performed/administered<br>• Recommendation that procedure or treatment is useful/effective<br>Class IIa<br>• Benefit > > Risk<br>• Additional studies with focused objective needed<br>• IT IS REASONABLE to perform procedure/administer treatment<br>• Recommendation in favor of treatment or procedure being useful/effective |
| Provide all other grades and definitions from the recommendation grading system | Class of Recommendation:<br>Class IIb<br>• Benefit ≥ Risk<br>• Additional studies with broad objectives needed; additional registry data would be helpful<br>• Procedure/Treatment MAY BE CONSIDERED<br>• Recommendation's usefulness/efficacy less well established<br>Class III – No benefit<br>• Procedure/Test – Not Helpful<br>• Treatment – No proven benefit<br>• Recommendation that procedure or treatment is not useful/effective and may be harmful<br>Class III – Harm<br>• Procedure/Test – Excess cost without benefit or harmful<br>• Treatment – Harmful to patients |

| | |
|---|---|
| | • Recommendation that procedure or treatment is not useful/effective and may be harmful |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | The ACC/AHA does not provide information on the systematic review conducted to support its guideline and the recommendations mentioned above. In lieu of the ACC/AHA systematic review, we reported on another systematic review of the evidence that supports the ACC/AHA recommendations in Table 3. |
| Estimates of benefit and consistency across studies | See Table 3. |
| What harms were identified? | See Table 3. |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | There have been no new studies that contradict the current body of evidence. |

Table 3. Systematic Review

| | |
|---|---|
| **Citation** | **Both guidelines used to support this measure cover a much wider topic area than just secondary prevention of myocardial infarction with persistent beta-blocker therapy treatment and do not discuss in detail the evidence review process for each recommendation supporting the persistence of beta-blocker treatment after heart attack measure. They do, however, provide a grade of evidence for each of the recommendations and cite systematic reviews supporting those recommendations. Therefore, we are using the evidence grades the guidelines provide and reference one seminal systematic review cited in the guidelines that summarizes the body of evidence supporting the recommendations.**<br><br>• **Freemantle N, Cleland J, Young P, Mason J, Harrison J. Beta blockade after myocardial infarction: systematic review and meta regression analysis. BMJ. 1999;318:1730–1737. http://www.ncbi.nlm.nih.gov/pubmed/10381708** |
| What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review? | The evidence for this measure focuses on the importance of beta blocker therapy in long-term secondary prevention of acute myocardial infarction (AMI). It is important to note that a systematic evidence review completed by Freemantle et al., in 1999 supports and is referenced by both STEMI and NSTEMI guidelines.<br><br>Freemantle et al. assessed the effectiveness of beta-blockers in longer-term secondary prevention of AMI using randomized controlled trials (RCTs). The review focused on RCTs that compared beta-blockers to placebo. |
| Grade assigned for the quality of the quoted evidence with definition of the grade | Per Freemantle et al., grades assigned for the quality of the evidence varied from Level A – Level C. See Table 1 and Table 2 for the grade assigned to each guideline recommendation.<br><br>Level of Evidence:<br>Level A<br>• Multiple populations evaluated*<br>• Data derived from multiple randomized clinical trials or meta-analyses<br>Level C<br>• Very limited populations evaluated*<br>• Only consensus opinion of experts, case studies, or standard of care |
| Provide all other grades and associated definitions of the evidence in the grading system | Level of Evidence:<br>Level B<br>• Limited populations evaluated*<br>• Data derived from a single randomized trial or nonrandomized studies |

| | |
|---|---|
| What is the time period covered by the body of evidence? | It should be noted that the body of evidence supporting the guideline recommendations is much broader and includes more recent evidence than the evidence used in the Freemantle et al. systematic review which includes studies published from 1966-1997. |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | • There are 31 long-term randomized controlled trials included in the systematic evidence review by Freemantle et al. (1999), which supports the STEMI and NSTEMI guideline recommendations regarding persistent beta-blocker treatment after heart attack.<br>• The Freemantle et al. systematic evidence review rated the quality of studies as reasonably high, with adequate follow-up achieved in many trials. |
| Estimates of benefit and consistency across studies | Considerable evidence supports the routine long-term use of beta blockers in patients who have had a myocardial infarction, with substantial benefits in terms of reduced mortality and morbidity.<br><br>Freemantle et al., use a random effects approach in long term trials for incidence of risk difference to estimate to normalized annual reduction in mortality across trials. This approach suggests an annual reduction of 1.2 deaths in 100 patients treated with beta-blockers after myocardial infarction; that is about 84 patients will require treatment for 1 year to avoid one death. A similar approach was used to estimate the effects of treatment on reinfarction, although only 21 of the 34 comparisons provided data on reinfarction, resulting in wider confidence intervals and the potential for reporting bias. This analysis suggests an annual reduction in reinfarction of 0.9 events every 100 (0.3 to 1.6); that is about 107 patients would require treatment of 1 year to avoid one non-fatal reinfarction. There was a 23% reduction in the odds of death in long term trials (95% confidence interval 15% to 31%). |
| What harms were identified? | The guidelines and systematic review provide extremely limited findings regarding harm associated with persistent beta blocker treatment after a heart attack. Freemantle et al., studied withdrawal from treatment for both active treatment and placebo groups. The trials reported that dizziness, depression, cold extremities, and fatigue were only marginally more common in the treatment than control groups. This supports the fact that the benefits of beta-blocker treatment significantly outweigh the minor treatment harms. |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | There have been many (>100) studies examining the use of beta blockers in patients who have had an MI since the publication of the systematic reviews used to generate the STEMI and NSTEMI guidelines. An article published in 2012 by Bangalore et al., confirms that beta-blockers remain the standard of care after a myocardial infarction. This study also references the findings from the Freemantle et al., systematic review used to support the recommendations for our measure. |

| | |
|---|---|
| | • Bangalore S, Steg G, Deedwania P, et al. β-Blocker Use and Clinical Outcomes in Stable Outpatients With and Without Coronary Artery Disease. *JAMA.* 2012;308(13):1340-1349. doi:10.1001/jama.2012.12559. |

*Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use.

_____

**1a.4 OTHER SOURCE OF EVIDENCE**

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

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

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.*

This measure addresses the appropriate clinical management of a person who has experienced an AMI. Persistent beta-blocker treatment after a heart attack reduces the risk of mortality, reduces the risk of severity of reinfarction, and improves the preservation of the left ventricular function.

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

The following data are extracted from HEDIS data collection and reflect the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by the mean, standard deviation, minimum health plan performance, maximum health plan performance, performance percentiles (10th, 25th, 50th, 75th, and 90th percentile) and the interquartile range. Data is stratified by year and product line (i.e. commercial, Medicare, Medicaid) at the health plan level.

Persistence of Beta-Blocker Treatment After a Heart Attack

N = Number of Health Plans

YEAR = Measurement Year

Commercial

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

2015|245|83%|6%|62%|76%|79%|83%|88%|91%|99%|9%

2016|251|84%|7%|57%|76%|80%|85%|89%|92%|98%|9%

2017|243|85%|6%|57%|77%|81%|85%|89%|92%|100%|8%

Medicaid

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

2015|115|80%|11%|43%|64%|75%|83%|88%|92%|97%|13%

2016|136|80%|9%|50%|67%|77%|81%|86%|90%|95%|9%

2017|145|78%|9%|39%|66%|74%|80%|84%|89%|97%|10%

Medicare

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

2015|258|91%|5%|68%|85%|88%|91%|94%|97%|100%|6%

2016|256|90%|5%|61%|83%|88%|91%|94%|96%|100%|6%

2017|272|90%|5%|71%|84%|88%|91%|93%|95%|100%|6%

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

N/A

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

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

The report described racial and ethnic disparities among beneficiaries 18 and older who received persistent beta blocker treatment for 6-months following a hospital discharge for a heart attack. Overall, Whites were more likely to receive treatment. Whites received treatment over 3% more than Blacks, at a rate of 92.2% while Blacks received treatment at 86.8%, respectively. Hispanic beneficiaries received treatment at a slightly higher rate than Blacks, at 87.6%, but still remain under treated compared to Whites. White beneficiaries were also more likely to receive treatment than Asian or Pacific Islanders, but well within 3 percentage points of each other. Pacific Islanders or Asians were treated at a rate of 90.0%.

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

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

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

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

Heart disease is the leading cause of death for people of most ethnicities in the United States, including African Americans, Hispanics, and whites. For American Indians or Alaska Natives and Asians or Pacific Islanders, heart disease is the second leading cause of death (CDC, 2017). Non-Hispanic black adults are at least 50% more likely to die of heart disease or stroke prematurely (i.e., before age 75 years) than their non-Hispanic white counterparts (CDC, 2013). Black women and men are more likely to die before age 75 as a result of coronary heart disease (CHD) than white women and men (rates of death are 37.9%, 61.5%, 19.4%, and 41.5%, respectively) (CDC, 2011). Racial and age-related disparities also exist in rates of recurrent MI or fatal CHD within 5 years of a first MI. Of those who have a first MI, the percentage with a recurrent event is as follows: at 45 to 64 years of age, 14% of white men, 18% of white women, 22% of black mean, and 28% of black women; at >=65 years of age, 21% of white men and women, 33% of black men, and 26% of black women (Mozaffarian et al., 2015).

A 2012 study by Zhang et al. compared medication adherence among MI survivors by disability, status, race/ethnicity, and income for all Medicare fee-for-service beneficiaries discharged post-MI in 2008. Among the disabled who were taking beta-blockers, the percentage of beneficiaries with good adherence for 6-month adherence was highest for Whites at 67% and lowest for Blacks at 52% with Asians, Hispanics, and Native Americans ranging in between (Zhang et al., 2012).

The CDC analyzed data from 2008-2012 to identify if employment status had an impact on rates of CHD/stroke. The results of this analysis showed that 1.9% of employed adults aged <55 years reported a history of CHD/stroke, compared with 2.5% of unemployed adults looking for work, and 6.3% of adults not in the labor force. Workers employed in service and blue-collar occupations were more likely than those in white collar occupations to report a history of CHD/stroke (Luckhaupt, 2014).

Center for Disease Control and Prevention (CDC). 2017. Heart Disease Facts. Last modified November 28, 2017. http://www.cdc.gov/heartdisease/facts.htm

Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services. 2013. "CDC Health Disparities and Inequalities Report-United States, 2013." Morbidity and Mortality Weekly Report (MMWR) 62(03); 1-2. http://www.cdc.gov/heartdisease/facts.htm

Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services. 2011. "Fact Sheet: Health Disparities in Coronary Heart Disease and Stroke." http://www.cdc.gov/minorityhealth/CHDIR/2011/FactSheets/CHDStroke.pdf

Luckhaupt, S.E., Calvert, G.M. August 2014. "Prevalence of Coronary Heart Disease or Stroke Among Workers Aged <55 years-United States 2008-2012." Morbidity and Mortality Weekly Report (MMWR). 63(30); 645-649. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6630a1.htm

Mozaffarian, D., Benjamin, E.J., Go, A.S., et al. 2015. "Heart Disease and Stroke Statistics-2015 Update: A Report from the American Heart Association." Circulation. 131:e29-e322. doi: 10.1161/CIR.0000000000000152

Zhang, Y., Baik, S.H., Chang, C-C.H., Kaplan, C.M., Lave, J.R. 2012. "Disability, Race/ethnicity, and Medication Adherence Among Medicare Myocardial Infarction Survivors." American Heart Journal. 164(3): 425-433.e4. doi: 10.1016/j.ahj.2012.05.021.

## 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):*

Cardiovascular, Cardiovascular : Coronary Artery Disease (AMI)

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

Primary Prevention

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

Populations at Risk

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

N/A

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** 0071_PBH_Value_Sets_Fall_2019-637091548789757231.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

There have been minor changes to the value sets and medication lists to reflect current practice.

NCQA added a hospice exclusion to HEDIS measures in 2016. The focus of hospice care is not to cure illnesses of patients, but rather to improve comfort and quality of life for those with limited life expectancy. Most HEDIS quality measures are focused on health screenings or treatments that are not clinically appropriate or beneficial for those who are at end of life. Many of these screenings and treatments would also be uncomfortable or pose risks for hospice patients, add undue burden and have no impact on improving length or quality of life. Therefore, including individuals who are receiving hospice in this measure is inappropriate.

In addition, NCQA added exclusion criteria for adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings. We recognize that for individuals with limited life expectancy, advanced illness or more complex clinical situations, the treatment identified in this measure may not be relevant or in line with the patient's goals of care. By implementing this set of exclusions, those providing care to the frail and advanced illness population can focus on care that's more appropriate for their conditions and health status. Attention can be more focused on quality measures that capture services and care processes that are most relevant for this population (e.g., improving care transitions, getting follow-up after acute care episodes, or avoiding preventable hospitalizations).

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

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

Patients who received at least 135 days of treatment with beta-blockers during the 180-day measurement interval.

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

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

At least 135 days of treatment with beta-blockers during the 180-day measurement interval.

180-day measurement interval – The 180-day period that includes the discharge date and the 179 days after discharge.

To determine continuity of treatment during the 180-day period, identify all prescriptions filled within the 180-day measurement interval, and add the number of allowed gap days (up to a total of 45 days) to the number of treatment days for a maximum of 180 days (i.e., 135 treatment days + 45 gap days = 180 days).

Treatment days (days covered) – The actual number of calendar days covered with prescriptions within the specified 180-day measurement interval (i.e., a prescription of a 90-day supply dispensed on the 100th day will have 80 days counted in the 180-day interval).

Assess for active prescriptions and include days supply that fall within the 180-day measurement interval. For patients who were on beta-blockers prior to admission and those who were dispensed an ambulatory

prescription during their inpatient stay, factor those prescriptions into adherence rates if the actual treatment days fall within the 180-day measurement interval.

PBH-B BETA-BLOCKER MEDICATIONS

DESCRIPTION / PRESCRIPTION

Noncardioselective beta-blockers / Carvedilol; Labetalol; Nadolo; Penbutolol; Pindolol; Propranolol; Timolol; Sotalol

Cardioselective beta-blockers / Acebutolol; Atenolol; Betaxolol; Bisoprolol; Metoprolol; Nebivolol

Antihypertensive combinations / Atenolol-chlorthalidone; Bendroflumethiazide-nadolol; Bisoprolol-hydrochlorothiazide; Hydrochlorothiazide-metoprolol; Hydrochlorothiazide-propranolol

See attached code value sets.

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

An acute inpatient discharge from July 1 of the year prior to the measurement year through June 30 of the measurement year with any diagnosis of acute myocardial infarction (AMI) on the discharge claim.

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

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

Patients who had continuous enrollment from discharge date through 179 days after discharge. No more than one gap in continuous enrollment of up to 45 days within the 180 days of the event. If the patient has Medicaid, then no more than a 1-month gap in coverage.

An acute inpatient discharge from July 1 of the year prior to the measurement year through June 30 of the measurement year with any diagnosis of acute myocardial infarction (AMI) on the discharge claim.

To identify an acute inpatient discharge:

1. Identify all acute and nonacute inpatient stays.

2. Exclude nonacute inpatient stays.

3. Identify the discharge date for the stay.

If a patient has more than one episode of AMI that meets the event/diagnosis criteria, from July 1 of the year prior to the measurement year through June 30 of the measurement year, include only the first discharge.

Direct transfers to an acute inpatient care setting: If a patient had a direct transfer to an acute inpatient setting (for any diagnosis), use the discharge date from the transfer setting, not the initial discharge. Exclude both the initial discharge and the direct transfer discharge if the transfer discharge occurs after June 30 of the measurement year. Use the instructions below to identify direct transfers and exclude nonacute inpatient stays.

Direct transfers to a nonacute inpatient care setting: Exclude from the denominator, hospitalizations in which the patient had a direct transfer to a nonacute inpatient care setting for any diagnosis. Use the instructions below to identify direct transfers and confirm the stay was for nonacute inpatient care based on the presence of a nonacute code on the claim.

A direct transfer is when the discharge date from the first inpatient setting precedes the admission date to a second inpatient setting by one calendar day or less. For example:

- An inpatient discharge on June 1, followed by an admission to another inpatient setting on June 1, is a direct transfer.

- An inpatient discharge on June 1, followed by an admission to an inpatient setting on June 2, is a direct transfer.

- An inpatient discharge on June 1, followed by an admission to another inpatient setting on June 3, is not a direct transfer; these are two distinct inpatient stays.

Use the following method to identify admissions to and discharges from inpatient settings.

1. Identify all acute and nonacute inpatient stays.

2. If needed, identify nonacute inpatient stays.

3. Identify the admission and discharge dates for the stay.

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

Any of the following any time during the patient's history through the end of the continuous enrollment period meet criteria:

- Asthma

- COPD

- Obstructive chronic bronchitis

- Chronic respiratory conditions due to fumes and vapors

- Hypotension, heart block >1 degree or sinus bradycardia

- A medication dispensing event indicative of a history of asthma

- Intolerance or allergy to beta-blocker therapy

Additionally, this measure excludes adults in hospice. It also excludes adults with advanced illness and frailty, as well as Medicare adults 65 years of age and older enrolled in an I-SNP or living long-term in institutional settings.

**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.)*

Patients identified as having an intolerance or allergy to beta-blocker therapy. Any of the following any time during the patient's history through the end of the continuous enrollment period meet criteria:

- Asthma

- COPD

- Obstructive chronic bronchitis

- Chronic respiratory conditions due to fumes and vapors

- Hypotension, heart block >1 degree or sinus bradycardia

- A medication dispensing event indicative of a history of asthma

MEDICATIONS TO IDENTIFY HISTORY OF ASTHMA

DESCRIPTION / PRESCRIPTION

Bronchodilator combinations / Budesonide-formoterol; Fluticasone-vilantero; Fluticasone-salmeterol; Formoterol-mometasone

Inhaled corticosteroids / Beclomethasone; Budesonide; Ciclesonide; Flunisolide; Fluticasone; Mometasone

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

Exclude adults who meet any of the following criteria:

- Medicare members 66 years of age and older as of December 31 of the measurement year who meet either of the following:

-- Enrolled in an Institutional SNP (I-SNP) any time on or between July 1 of the year prior to the measurement year and the end of the measurement year.

-- Living long-term in an institution any time on or between July 1 of the year prior to the measurement year and the end of the measurement year as identified by the LTI flag in the Monthly Membership Detail Data File. Use the run date of the file to determine if an adult had an LTI flag any time on or between July 1 of the year prior to the measurement year and the end of the measurement year.

- Members 66-80 years of age as of December 31 of the measurement year (all product lines) with frailty and advanced illness. Adults must meet BOTH of the following frailty and advanced illness criteria to be excluded:

1. At least one claim/encounter for frailty any time on or between July 1 of the year prior to the measurement year and the end of the measurement year.

2. Any of the following during the measurement year or the year prior to the measurement year (count services that occur over both years):

-- At least two outpatient visits, observation visits, ED visits, nonacute inpatient encounters or nonacute inpatient discharges (instructions below) on different dates of service, with an advanced illness diagnosis. Visit type need not be the same for the two visits. To identify a nonacute inpatient discharge:

1. Identify all acute and nonacute inpatient stays.

2. Confirm the stay was for nonacute care based on the presence of a nonacute code on the claim.

3. Identify the discharge date for the stay.

-- At least one acute inpatient encounter with an advanced illness diagnosis.

-- At least one acute inpatient discharge with an advanced illness diagnosis. To identify an acute inpatient discharge:

1. Identify all acute and nonacute inpatient stays.

2. Exclude nonacute inpatient stays.

3. Identify the discharge date for the stay.

-- A dispensed dementia medication.

DEMENTIA MEDICATIONS

DESCRIPTION / PRESCRIPTION

Cholinesterase inhibitors / Donepezil; Galantamine; Rivastigmine

Miscellaneous central nervous system agents / Memantine

- Members 81 years of age and older as of December 31 of the measurement year (all product lines) with frailty any time on or between July 1 of the year prior to the measurement year and the end of the measurement year.

See attached code value sets.

**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.)*

No stratification

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

No risk adjustment or risk stratification

If other:

**S.12. Type of score:**

Rate/proportion

If other:

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

Better quality = Higher score

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

STEP 1: Determine the eligible population. To do so, identify patients who meet all specified criteria.

- AGES: 18 years and older as of December 31 of the measurement year.

- EVENT/DIAGNOSIS: Identify patients who were discharged from an acute setting with an AMI from July 1 of the year prior to the measurement year through June 30 of the measurement year. SEE S.6 and S.7 for eligible population and denominator criteria and details.

STEP 2: Exclude patients who meet the exclusions criteria. SEE S.8 and S.9 for denominator exclusion criteria and details.

STEP 3: Determine the number of patients in the eligible population who were given a 180-day course of treatment with beta blockers post discharge.

STEP 4: Identify patients whose dispensed days´ supply is >=135 days in the 180-day measurement interval. SEE S.4 and S.5 for numerator criteria and details.

STEP 5: Calculate the rate by dividing the numerator (STEP 4) by the denominator (after exclusions) (STEP 2).

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

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

N/A

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

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

N/A

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

*If other, please describe in S.18.*

Claims

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

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

This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from health plans via NCQA's online data submission system.

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

No data collection instrument provided

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

Health Plan

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

Outpatient Services

If other:

**S.22. <u>COMPOSITE Performance Measure</u>** - 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**

PBH_Testing_Form_11.20.2019-637099345185494109.docx

**2.1 <u>For maintenance of endorsement</u>**

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

Yes

**2.2 <u>For maintenance of endorsement</u>**

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

Yes

**2.3 <u>For maintenance of endorsement</u>**

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

No - This measure is not risk-adjusted

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

**Measure Number** (*if previously endorsed*)**:** 0071
**Measure Title**: Persistence of Beta-Blocker Treatment After a Heart Attack (PBH)
**Date of Submission**: 8/1/2019

**Type of Measure:**

| ☐ **Outcome (*including PRO-PM*)** | ☐ **Composite – *STOP – use composite testing form*** |
|---|---|
| ☒ Intermediate Clinical Outcome | ☐ Cost/resource |

| | |
|---|---|
| ☐ Process *(including Appropriate Use)* | ☐ Efficiency |
| ☐ Structure | |

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

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

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

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

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

N/A

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

Testing of performance measure score with beta binomial reliability and testing of construct validity with the Pearson Correlation were performed using HEDIS 2018 plan level data, measurement year 2017.

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

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

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

This measure assesses whether adults enrolled in commercial, Medicare, and Medicaid plans who had acute myocardial infarctions received persistent beta blocker treatment six months after discharge. Therefore, testing was done at the health-plan level, which is appropriate for the level of reporting for this measure.

We calculated the measure score reliability and construct validity from HEDIS data that included 243 commercial plans, 145 Medicaid plans, and 272 Medicare plans. The sample included all commercial, Medicare, and Medicaid health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

**1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
Below is a description of the data submitted for 2017, including the median denominator size per plan. Data are summarized at the health plan level and stratified by plan type (i.e. commercial, Medicaid, Medicare).

Table 1. Median denominator size per plan for Persistence of Beta-Blocker Treatment After a Heart Attack, 2017.

| Product Line | Number of Plans | Median Denominator Size/Plan |
|---|---|---|
| Commercial | 243 | 65 |
| Medicaid | 145 | 81 |
| Medicare | 272 | 72 |

**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**.

Reliability:

Reliability of the health plan measure score was tested using a beta-binomial calculation. This analysis included the entire HEDIS data sample (described above).

Validity:

Validity of the health plan measure was demonstrated through construct validity using the entire HEDIS data sample (described above).

**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 did not analyze social risk factors. This measure of health plan performance is specified to be reported separately by commercial, Medicaid, and Medicare plan types, which serves as a proxy for income and other socioeconomic factors.

**2a2. RELIABILITY TESTING**

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

**2a2.1. What level of reliability testing was conducted**? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

Reliability was estimated by using the Beta-binomial model (Adams, 2009) for this health plan measure. Beta-binomial is appropriate for estimating the reliability of pass/fail rate measures. Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

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

Table 2 provides the reliability for the overall measure as shown by the Beta-binomial model and the distribution of individual plan reliability.

Table 2. Overall Beta-binomial statistic and distribution of plan reliability for commercial, Medicaid, and Medicare product lines, 2017

| Product Line | Overall Reliability | Min | Percentile | | | | | Max |
|---|---|---|---|---|---|---|---|---|
| | | | 10th | 25th | 50th | 75th | 90th | |
| Commercial | 0.757 | 0.247 | 0.396 | 0.521 | 0.672 | 0.808 | 0.876 | 1.000 |
| Medicaid | 0.818 | 0.149 | 0.264 | 0.389 | 0.621 | 0.800 | 0.889 | 1.000 |
| Medicare | 0.739 | 0.402 | 0.554 | 0.670 | 0.772 | 0.861 | 0.919 | 0.976 |

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

The values for the overall beta-binomial statistic across all product lines for the measure are all greater than 0.7, indicating the measure has very good reliability. The distribution of health plan level-reliability on this measure shows that approximately half of health plans (across all product lines) are either right at the threshold of 0.7 or exceed it. Good reliability is demonstrated since most variance is due to signal and not to noise.

**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)*

We tested for construct validity of the Persistence of Beta-Blocker Treatment After a Heart Attack (PBH) measure by exploring whether it was correlated with another similar measure of quality which is described below.

> Statin Therapy for Patients with Cardiovascular Disease (SPC) - Adherence 80%: The percentage of adults with clinical atherosclerotic cardiovascular disease who received a statin medication and achieved an 80% adherence threshold during the treatment period.

This measure was chosen for construct validity testing because it is similarly focused on a population with cardiovascular disease and includes an assessment of medication adherence during a specified timeframe. We hypothesized that, irrespective of an event-based or diagnosis-based measure, a health plan that does well on the statin adherence measure for cardiovascular disease would also do well on a measure of beta blocker persistence for patients who have had a heart attack.

To test this correlation, we used a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable. Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

** Note: All HEDIS value sets are updated annually with the most current codes available. The information below details the process we used to convert value sets that used ICD-9 codes to ICD-10 codes in 2015. **

ICD-10 Conversion:

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

Steps in ICD-9 to ICD-10 Conversion Process

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

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

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

**2b1.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test*)
The results from construct validity testing of the health plan level measure are presented by product line in Tables 3a, 3b, and 3c below.

Table 3a. Correlations between PBH and SPC-Adherence measures in Commercial Health Plans, 2017.

| | **Pearson Correlation Coefficients** |
|---|---|
| | SPC-Adherence |
| PBH | 0.51 |

Note: All correlations are significant at p<0.0001

Table 3b. Correlations between PBH and SPC-Adherence measures in Medicaid Health Plans, 2017.

| | **Pearson Correlation Coefficients** |
|---|---|
| | SPC-Adherence |
| PBH | 0.60 |

Note: All correlations are significant at p<0.0001

Table 3c. Correlations between PBH and SPC-Adherence measures in Medicare Health Plans, 2017.

| | **Pearson Correlation Coefficients** |
|---|---|

| | SPC-Adherence |
|---|---|
| PBH | 0.42 |

Note: All correlations are significant at p<0.0001

**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?*)

Across all product lines, the correlations are moderate and statistically significant, which suggests plan performance on the Statin Therapy for Patients with Cardiovascular Disease - Adherence 80% measure is correlated to performance on the Persistence of Beta-Blocker Treatment After a Heart Attack measure. Plans that have higher rates on one measure will have higher rates on the other. Coefficients with absolute value of less than .3 are generally considered indicative of weak associations. Absolute values of .3 to .59 are considered moderate associations, absolute values of .6 to .69 indicate a strong positive relationship, and absolute values of .7 or higher indicate a very strong positive relationship. These correlation results suggest that at the plan level the measure has sufficient validity.

_____

**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*)

We did not perform testing of the following exclusions for this submission:

- Asthma, as well as medication dispensing events indicative of a history of asthma
- COPD
- Obstructive chronic bronchitis
- Chronic respiratory conditions due to fumes and vapors
- Hypotension, heart block >1 degree or sinus bradycardia
- Intolerance or allergy to beta-blocker therapy

NCQA engaged expert panels to inform the face validity of these exclusions, which align with the evidence and guideline recommendations supporting the measure. This measure has been reviewed by NCQA's Cardiovascular Measurement Advisory Panel, Technical Measurement Advisory Panel, and the Committee on Performance Measurement. The measure also received public comment feedback upon initial development.

Hospice, I-SNPs and Long-Term Care Institutions

These exclusions were also not formally tested for this submission. This measure is designed to be scientifically valid and feasible for comparing the quality of care provided to general populations, such as healthy older adults or those with a single condition. Patients receiving hospice, enrolled in an I-SNP, or residing in a long-term care institution would likely have different care needs and quality concerns, therefore they are excluded from this measure.

Advanced Illness and Frailty

For HEDIS 2019 (measurement year 2018), NCQA added exclusions for advanced illness and frailty to the Persistence of Beta-Blocker Treatment after a Heart Attack measure. NCQA decided to explore implementing these exclusions, recognizing that for individuals with limited life expectancy, advanced illness or frailty, the treatment identified in this measure may not be clinically appropriate, relevant or in line with the patient's

goals of care. We performed a review of literature on different approaches to defining advanced illness and used this, along with feedback received from expert work groups, measurement advisory panels and public comment to create a list of illnesses, conditions and service codes to be included in testing. The conditions included: dementia and other neurodegenerative conditions, emphysema, end stage renal disease (ESRD), heart failure, liver failure, metastatic cancer, pulmonary fibrosis and respiratory failure.

NCQA then conducted a search of ICD-10 codes that were relevant to each of the conditions to create value sets for testing. To identify those with dementia, NCQA also included drug codes for medications such as donepezil hydrochloride and galantamine hydrobromide, to capture those who may not carry a diagnosis of dementia but are prescribed a drug for treatment.

The proxy for frailty was developed based on previously studied approaches[1,2,3] and feedback received from expert work groups and measurement advisory panels. The proxy is comprised of HCPCS and ICD-10 codes for diagnoses or services that can indicate when an individual is frail or dependent in activities of daily living. Examples include: gait abnormality, abnormal loss of weight and underweight, adult failure to thrive, debility, fall, pressure ulcer, durable medical equipment (hospital bed, walker, portable or home oxygen, wheelchair), bed confinement, palliative care and age-related physical debility. Members met the frailty proxy criteria if they had a claim for any of the codes included in the frailty code set in the measurement year.

To determine the feasibility and impact of applying these exclusions to the measure, NCQA used a research database that consisted of two years of inpatient, outpatient, and pharmacy claims for members age 18 and older enrolled in a sample of Medicare Advantage plans (N=10). NCQA compared several approaches for identifying the advanced illness and frailty populations, examining different age ranges and diagnosis positions and their impact on the denominator size and performance rate of the measure. The results of those queries along with input from the expert work groups, measurement advisory panels and public comment led us to determine that the best approach for identifying the advanced illness and frailty population that should be excluded from the measure was to apply the following criteria:

- Adults 66–80 years of age as of December 31 of the measurement year (all product lines) with frailty and advanced illness
- Adults 81 years of age and older as of December 31 of the measurement year with frailty any time on or between July 1 of the year prior to the measurement year and the end of the measurement year.

**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*)
Table 4a shows the results of applying the exclusion of adults 66–80 years of age with frailty and advanced illness to the Persistence of Beta-Blocker Treatment after a Heart Attack (PBH) measure. Table 4b shows the results of applying the exclusion of adults 81 and older with frailty.

---

[1] Faurot, K.R., Funk, M.J., Pate, V., Brookhart, M.A., Patrick, A., Hanson, L.C., Castillo, W.C., Stürmer, T. 2015. Using Claims Data to Predict Dependency in Activities of Daily Living as a Proxy for Frailty. Pharmacoepidemiology and Drug Safety. 24(1): 59-66.

[2] Segal, J.B., Chang, H.Y., Du, Y., Walston, J.D., Carlson, M.C., Varadhan, R. 2017. Development of a

Claims-Based Frailty Indicator Anchored to a Well-Established Frailty Phenotype. Medical Care. 55(7): 716-722.

[3] Davidoff A.J., A. Hurrida, I.H. Zuckerman, S.M. Lichtman, N. Pandya, A. Hussain, F. Hendrick, J.P. Weiner, X. Ke, M.J. Edelman. 2013. A Novel Approach to Improve Health Status Measurement in Observational Claims-Based Studies of Cancer Treatment and Outcomes. J Geriatr Oncol. 4(2):157–165.

Table 4a. Impact of applying the advanced illness and frailty for patients aged 66-80 exclusion to the PBH measure

| Number of Plans (N) | Average Number Excluded | Average % Removed by Exclusion | Average Performance Rate without Exclusion (%) | Average Performance Rate with Exclusion (%) | Difference in Average Rate (%) |
|---|---|---|---|---|---|
| 10 | 13 | 4.6 | 57.1 | 59.6 | 2.5 |

Table 4b. Impact of applying the frailty exclusion for patients 81 and older to the PBH measure

| Number of Plans (N) | Average Number Excluded | Average % Removed by Exclusion | Average Performance Rate without Exclusion (%) | Average Performance Rate with Exclusion (%) | Difference in Average Rate (%) |
|---|---|---|---|---|---|
| 10 | 3 | 1.1 | 57.1 | 57.6 | 0.5 |

**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*)

Advanced Illness and Frailty

The advanced illness and frailty exclusion had a small impact on the eligible population: 4.6% on average were removed for advanced illness; 1.1% on average were removed for frailty. Impact on performance rates was minimal. Feedback from NCQA's expert work groups and measurement advisory panels, as well as public comment feedback, supported the application of these exclusions to the Persistence of Beta-Blocker Treatment after a Heart Attack measure for clinical reasons. By implementing this set of exclusions, those providing care to the frail and advanced illness population can focus on care that is more appropriate for their conditions and health status. Attention can be more focused on quality measures that capture services and care processes that are most relevant for this population (e.g., improving care transitions, getting follow-up after acute care episodes, or avoiding preventable hospitalizations).

---

**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** Click here to enter number of factors **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.**

N/A

**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**.

The Persistence of Beta-Blocker Treatment after a Heart Attack measure assesses whether health plan members were dispensed beta blockers over the six months following an acute myocardial infarction (AMI). The measure is not assessing an outcome such as AMI-related morbidity or mortality for which a clinical factor may affect a health plan's ability to ensure medications are dispensed. It assesses persistence of beta-blocker treatment in a defined period of time for a population where there is strong evidence to support the benefit of the medication. Because the conceptual basis for risk adjustment of adherence measures is still developing, NCQA does not currently risk adjust this measure given the potential to mask poor performance and disparities in care in patients for whom evidence supports prescribing beta-blocker treatment following AMI.

**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?

N/A

**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.**

N/A

**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*)

N/A

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

***If stratified, skip to 2b3.9***

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

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

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

**2b3.9. Results of Risk Stratification Analysis**:

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

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

---

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

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

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure.

To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p-value of the test statistic is less than .05, then the two plans' performance is significantly different from each other.

**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*)

Table 5. Variation in Performance for Commercial, Medicaid, and Medicare health plans, 2017.

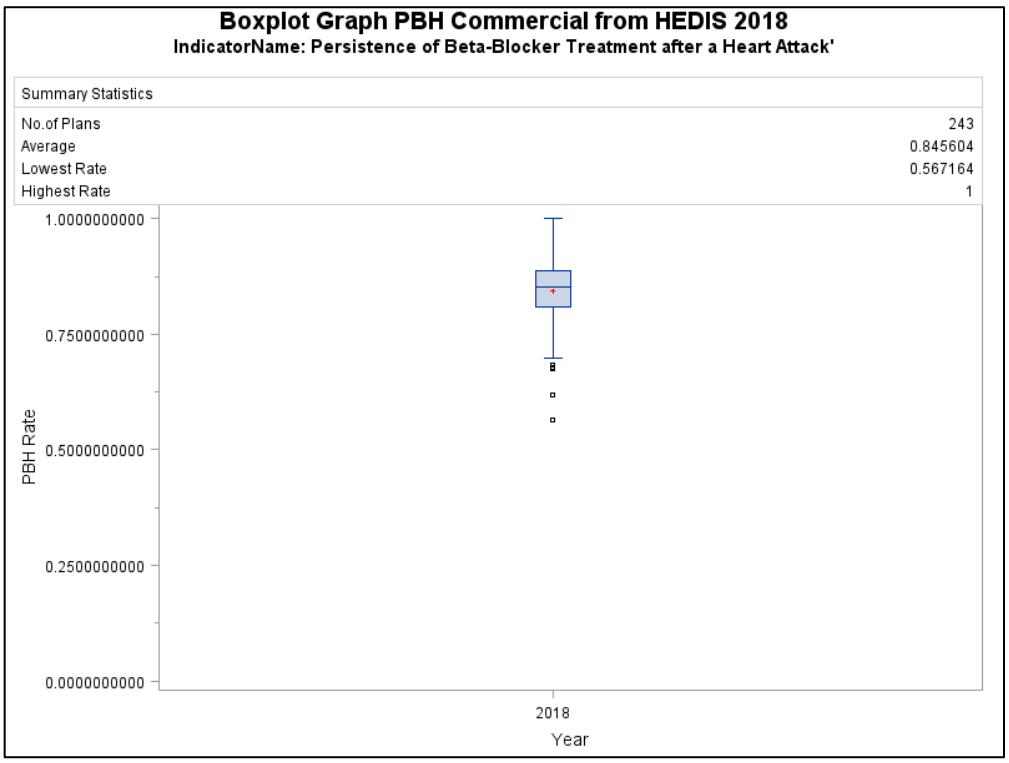
| Plan Type | N | Average (%) | St Dev (%) | 10th (%) | 25th (%) | 50th (%) | 75th (%) | 90th (%) | IQR (%) | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| Commercial | 243 | 84.56 | 6.36 | 76.58 | 80.83 | 85.31 | 88.68 | 91.89 | 7.85 | <0.05 |
| Medicaid | 145 | 78.46 | 8.97 | 66.18 | 73.81 | 79.67 | 83.87 | 88.84 | 10.06 | <0.05 |
| Medicare | 272 | 90.15 | 4.66 | 83.72 | 87.71 | 90.50 | 93.43 | 95.41 | 5.72 | <0.05 |

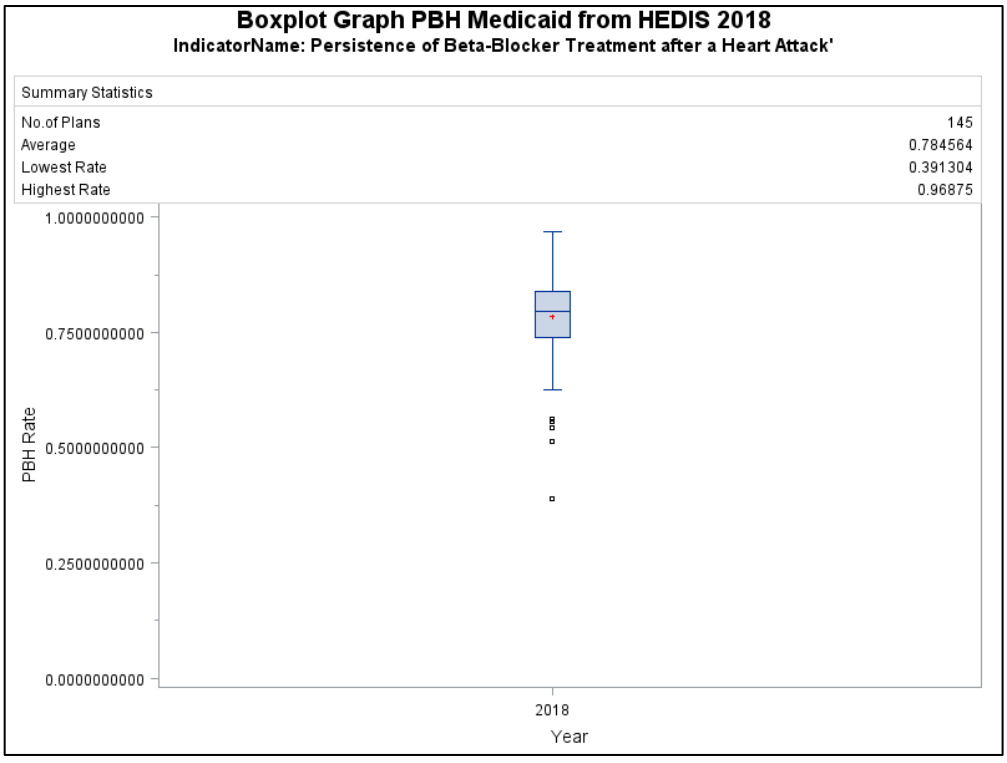
N = Number of plans reporting

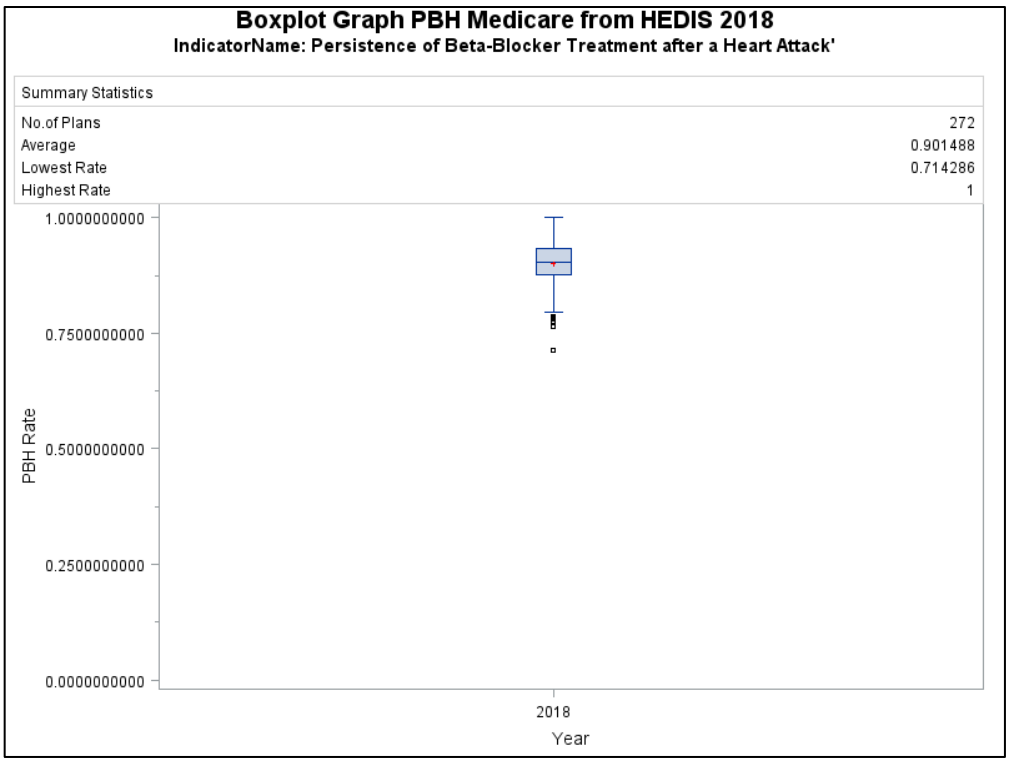
IQR = Interquartile range

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

Box plots for HEDIS 2018 (Measurement year 2017) Variation in Performance Across Health Plans are included below for your reference.






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

The results above indicate there is meaningful difference in performance. Across Medicaid, commercial and Medicare plans, the difference between the 25th and 75th percentile (better performance) is statistically significant.

---

**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.***

The Persistence of Beta-Blocker Treatment After a Heart Attack measure has only one set of specifications.

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

N/A

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

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

---

**2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

**2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
HEDIS measures apply to enrolled members in a health plan, and NCQA has a rigorous audit process to ensure the eligible population, denominator, and numerator events for each measure are correctly identified and reported. The audit process is designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.

The HEDIS Compliance Audit addresses the following functions:
- Information practices and control procedures
- Sampling methods and procedures
- Data integrity
- Compliance with HEDIS specifications
- Analytic file production
- Reporting and documentation

**2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data?** (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used. Once measures are added to HEDIS, NCQA conducts a first-year analysis to assess the measure's feasibility once widely implemented in the field. This analysis includes an assessment of how many plans report valid rates vs. rates that are materially biased (or have other issues, such as small denominators). These considerations are weighed in the deliberation process before measures are approved for public reporting.

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

This measure goes through the NCQA audit process each year to identify potential errors or bias in results. Only performances rates that have been reviewed and determined not to be "materially biased" are reported and used.

## 3. Feasibility

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

**3a. Byproduct of Care Processes**

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

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

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

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.**

NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO's ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

1) Information practices and control procedures
2) Sampling methods and procedures
3) Data integrity
4) Compliance with HEDIS specifications
5) Analytic file production
6) Reporting and documentation

In addition to the HEDIS audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system, NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system informs both annual updates to the measures as well as routine re-evaluation of measures. These processes include updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence.

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

Broad public use and dissemination of this measure is encouraged. NCQA has agreed with NQF that noncommercial users do not require the consent of the measure developer. Use by health care providers in connections with their own practices is not commercial use. Commercial use of a measure requires the period written consent of NCQA. As used herein, "commercial use" refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

## 4. Usability and Use

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

**4a. Accountability and Transparency**

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

**4.1. Current and Planned Use**

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

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|

| | |
|---|---|
| | Public Reporting<br>Health Plan Ratings<br>Report Cards<br>https://www.ncqa.org/hedis/reports-and-research/ratings-2019/<br>https://reportcards.ncqa.org/#/health-plans/list<br>Health Plan Ratings<br>Report Cards<br>https://www.ncqa.org/hedis/reports-and-research/ratings-2019/<br>https://reportcards.ncqa.org/#/health-plans/list<br>Regulatory and Accreditation Programs<br>NCQA Accreditation<br>https://www.ncqa.org/programs/health-plans/health-plan-accreditation-hpa/<br>NCQA Accreditation<br>https://www.ncqa.org/programs/health-plans/health-plan-accreditation-hpa/<br>Quality Improvement (external benchmarking to organizations)<br>Quality Compass<br>http://www.ncqa.org/hedis-quality-measurement/quality-measurement-products/quality-compass<br>Annual State of Health Care Quality<br>https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report/ |

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

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

STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2018, the report included results from calendar year 2017 for health plans covering a record 136 million people, or 43 percent of the U.S. population.
HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan rankings which are reported in Consumer Reports and on the NCQA website. These rankings are based on performance on HEDIS measures among other factors. In 2019, a total of 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Heath Plans. As of Fall 2017, a total of 184 Medicare Advantage health plans were accredited using this measure among others covering 9.2 million Medicare beneficiaries; 451 commercial health plans covering 113 million lives; and 125 Medicaid health plans covering 35 million lives. Health plans are scored based on performance compared to benchmarks.

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

N/A

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

N/A

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

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

Health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA. NCQA publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.

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

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Update and Best Practices Conference (now the Health Care Quality Congress), NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly and insight into new measure development projects. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c.1.

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

**Describe how feedback was obtained.**

NCQA measures are evaluated regularly using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

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

Questions received through the Policy Clarification Support System have generally centered around clarifications in the specification language, suggestions for potential exclusions, and clarifications on the recently added exclusion for advanced illness and frailty.

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

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

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

We have provided minor clarifications about the measure during the annual update process in order to address questions received through the Policy Clarification Support System.

**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.**

Over the past three years, Commercial plan performance has increased each year by about 1%; Medicare plan performance has remained relatively stable; a slight decrease in Medicaid plan performance was observed (2%). Current average performance (MY 2017) is highest in Medicare plans (90%), followed by commercial plans (85%), and then Medicaid plans (78%). We are encouraged by the sustained high performance across health plans but there is still room for improvement.

**4b2. Unintended Consequences**

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

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

There were no identified unexpected findings during testing or since implementation of this measure.

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

There were no identified unexpected findings during testing or since implementation of this measure.

## 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)**

0070 : Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)

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

**5a. Harmonization of Related Measures**

The measure specifications are harmonized with related measures;

**OR**

The differences in specifications are justified

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

**Are the measure specifications harmonized to the extent possible?**

No

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

DUE TO THE TEXT LIMIT IN THIS SECTION – WE ARE PROVIDING OUR ANSWER FOR 5a.2 IN SECTION 5b.1

**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.)**

ANSWER FOR SECTION 5a.2

NCQA's current Persistence of Beta Blocker Treatment After a Heart Attack measure (NQF measure 0071) uses health plan-reported data to assess the percentage of patients 18 years of age and older during the measurement year who were discharged with a diagnosis of AMI during the 6 months prior to the beginning of the measurement year through the 6 months after the beginning of the measurement year and who received persistent beta-blocker treatment for six months after discharge.

RELATED NQF MEASURE 0070 (Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)):

This measure assesses the percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12-month period who also have a prior MI or a current left ventricular ejection fraction (LVEF) <40% who were prescribed beta-blocker therapy.

HARMONIZED MEASURE ELEMENTS:

Measure 0071 and 0070 focus on patients 18 years and older who are prescribed beta-blocker treatment post-discharge after having a MI or history of MI. The National Quality Strategy Priorities classification for both measures is Prevention and Treatment of Cardiovascular Disease. Both measures exclude patients who are allergic or have an intolerance to beta blockers.

DIFFERENCES:

Below are the unharmonized measure elements between measure 0071 and measure 0070:

Measure 0071 focuses on beta-blocker treatment post a MI and Measure 0070 focuses on patients who have a prior MI or a current or prior LVEF <40%.

- Data Source: Data for measure 0071 is collected through administrative claims, electronic clinical data, and pharmacy data, while data for measure 0070 is collected through medical record, electronic health record data, electronic clinical data, and paper records
- Level of Accountability: Measure 0071 is a health plan level measure while measure 0070 is a clinician-level measure.
- Population: Measure 0071 focuses on patients who were diagnosed with a MI and discharged and prescribed a beta-blocker therapy treatment. Measure 0070 focuses on patients in a measurement year with a diagnosis of coronary artery diseases who also have a prior MI or current or prior LVEF.
- Exclusions: The difference in exclusions is that measure 0071 specifies asthma, COPD, obstructive chronic bronchitis, chronic respiratory conditions due to fumes and vapors, hypotension, hear block >1 degree, sinus bradycardia, and medication dispensing events indicative of a history of asthma as exclusions. Additionally, measure 0071 excludes hospitalizations in which the patient was transferred directly to a nonacute care facility

for any diagnosis, patients enrolled in an I-SNP, patients living long-term in an institution, patients 66-80 years of age with frailty and advanced illness, and patients 81 years of age and older with frailty. Measure 0070 exclusions include: documentation of patient reason(s) for not prescribing beta-blocker therapy (e.g., patient declined, other patient reasons) and documentation of system reason(s) for not prescribing beta-blocker therapy (e.g., other reasons attributable to the health care system).
IMPACT ON INTERPRETABILITY AND DATA COLLECTION BURDEN:
The differences between measures 0071 and 0070 do not have an impact on interpretability of publicly reported rates, or the burden of data collection, because all data for both measures are collected from different data sources by different entities.
ANSWER FOR SECTION 5b.1
Our current measure has a long-standing history of use by health plans and has been implemented for nearly 15 years.

## Appendix

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

No appendix **Attachment:**

## Contact Information

**Co.1 Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance

**Co.2 Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-

**Co.3 Measure Developer if different from Measure Steward:** National Committee for Quality Assurance

**Co.4 Point of Contact:** Kristen, Swift, nqf@ncqa.org, 202-955-1728-

## Additional Information

**Ad.1 Workgroup/Expert Panel involved in measure development**

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

NCQA follows a standard process of vetting members of the measurement advisory panel for conflicts of interest.

CARDIOVASCULAR MEASUREMENT ADVISORY PANEL

Kathy Berra, RN, MSN, ANP-BC, FAHA, FAAN, FPCNA, The LifeCare Company

Donald Casey, MD, MPH, MBA, FACP, FAHA, FAAPL, DFACMQ, American College of Medical Quality

Tom Kottke, MD, MSPH, HealthPartners

Eduardo Ortiz, MD, MPH, Tennessee Valley Healthcare System

Stephen Persell (Chair), MD, MPH, Northwestern University

Randall Stafford, MD, PhD, Stanford University

Kim Williams, MD, MACC, MASNC, FAHA, FESC, Rush University Medical Center

Tracy Wolff, MD, Agency for Healthcare Research and Quality

TECHNICAL MEASUREMENT ADVISORY PANEL

Andy Amster, MSPH, Kaiser Permanente

Sarah Bezeredi, MBA, MSHL, UnitedHealth Group

Jennifer Brudnicki, MBA, Inovalon Inc.

Lindsay Cogan, MS, PhD, New York State Department of Health

Mike Farina, MBA, R.Ph, Capital District Physicians' Health Plan

Marissa Finn, MBA, CIGNA

Scott Fox, MS, Med, FAMIA, The MITRE Corporation

Carlos Hernandez, CenCal Health

Harmon Jordan, ScD, Westat

Gigi Raney, LCSW, Center for Medicaid and CHIP Services

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

Laurie Spoll, Aetna

COMMITTEE ON PERFORMANCE MEASUREMENT

Andrew Baskin, MD, Aetna

Elizabeth Drye, MD, SM, Yale School of Medicine

Andrea Gelzer, MD, MS, FACP, AmeriHealth Caritas

Kate Goodrich, MD, MHS, Centers for Medicare & Medicaid Services

David Grossman, MD, MPH, Washington Permanente Medical Group

Christine Hunter, (Co-Chair), MD, WPS Health Solutions

David Kelley, MD, MPA, Pennsylvania Department of Human Services

Jeffrey Kelman, MMSc, MD, Department of Health and Human Services

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, Freelance

Adrienne Mims, MD, MPH, AGSF, FAAFP, Alliant Health Solutions

Amanda Parsons, MD, MBA, Metroplus

Wayne Rawlins, MD, MBA, ConnectiCare

Misty Roberts, MSN, RN, CPHQ, PMP, Humana

Rudy Saenz, MD, MMM, FACOG, Riverside Medical Clinic

Marcus Thygeson, (Co-Chair), MD, MPH, Blind On-Demand

JoAnn Volk, MA, Georgetown University

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 2005

**Ad.3 Month and Year of most recent revision:** 07, 2018

**Ad.4 What is your frequency for review/update of this measure?** Approximately every 3 years, sooner if the clinical guidelines have changed significantly

**Ad.5 When is the next scheduled review/update for this measure?** 12, 2020

**Ad.6 Copyright statement:** © 1999 by the National Committee for Quality Assurance

1100 13th Street, NW, 3rd Floor

Washington, DC 20005

**Ad.7 Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care and have not been tested for all potential applications. THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

**Ad.8 Additional Information/Comments:** Publication of each Measure is to be accompanied by the following notice:

NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA. © 2018 by the National Committee for Quality Assurance

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 #:** 0670

**Corresponding Measures:**

**De.2. Measure Title:** Cardiac stress imaging not meeting appropriate use criteria: Preoperative evaluation in low risk surgery patients

**Co.1.1. Measure Steward:** American College of Cardiology

**De.3. Brief Description of Measure:** Percentage of stress SPECT MPI, stress echo, CCTA, or CMR performed in low risk surgery patients for preoperative evaluation

**1b.1. Developer Rationale:** Appropriate use criteria define "when to do" and "how often to do" a given procedure in the context of scientific evidence, the health care environment, the patient's profile and a physician's judgment. While practice guidelines provide a foundation for summarizing evidence-based cardiovascular care or for providing expert consensus opinions, in many areas, marked variability remains in the use of cardiovascular procedures, raising questions about over-use and under-use. Appropriate use criteria provide practical tools to measure this variability and to look at utilization patterns. The criteria are designed to examine the use of diagnostic and therapeutic procedures to support efficient use of medical resources, while also providing patients with quality, appropriate care.

A measure that reports rates of inappropriate imaging within practices would contain information regarding both cost and quality, because an inappropriate test results in both higher costs and poorer-quality care. Conversely, a reduction in this rate would simultaneously improve quality and decrease cost. Improvements in this metric should lead to consistent application of AUC and improve the efficiency of the system.

**S.4. Numerator Statement:** Number of stress SPECT MPI, stress echo, CCTA, or CMR performed in patients undergoing low risk surgery as a part of the preoperative evaluation

**S.6. Denominator Statement:** Number of stress SPECT MPI, stress echo, CCTA, and CMR performed

**S.8. Denominator Exclusions:** None.

**De.1. Measure Type:** Efficiency

**S.17. Data Source:** Other, Registry Data

**S.20. Level of Analysis:** Clinician : Group/Practice, Facility

**IF Endorsement Maintenance – Original Endorsement Date:** Apr 26, 2011 **Most Recent Endorsement Date:** Jun 29, 2015

## Preliminary Analysis: Maintenance of Endorsement

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

## Criteria 1: Importance to Measure and Report

### 1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

**Evidence Summary of prior review in 2015**

- The developer provided evidence from the *2014 ACC/AHA Guideline on Perioperative Cardiovascular Evaluation and Management of Patients Undergoing Noncardiac Surgery*. The guidelines note that "Routine screening with noninvasive stress testing is not useful for patients undergoing low-risk noncardiac surgery."
    - The evidence was assigned "B" grade indicating "data derived from a single randomized trial, or non-randomized studies." The recommendation was assigned "Class III: No Benefit" grading, which corresponds to "conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective, and in some cases may be harmful."
- The developer notes that "only a few of the studies addressed the surgical population focused on in this measure." The studies are generally focusing on higher-risk surgeries than the low-risk surgeries that are a focus of this measure. The developer states it is reasonable to extrapolate the findings on higher-risk surgeries to low-risk surgeries.
- During the last endorsement review, the Committee expressed concerns that the developer does not provide evidence specific to appropriate use of imaging pre-operatively in low risk surgery patients.

**Changes to evidence from last review**

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

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

***Questions for the Committee:***

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

- What is the relationship of this measure to patient outcomes?
- How strong is the evidence for this relationship?
- Is the evidence directly applicable to the process of care being measured?

**Guidance from the Evidence Algorithm**

Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: moderate; Quality: moderate; Consistency: moderate (Box 5) → Moderate (Box 5b) → Moderate

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

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

**Maintenance measures – increased emphasis on gap and variation**

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

- For maintenance measures, performance scores on the measure as specified at the specified levels of analysis are required for maintenance of endorsement. The results provided do not appear to be calculated using the measure as specified. They do not include the same tests as the measure, and it is difficult to tell if the calculations were performed in alignment with the measure specifications.
- The developer presented site-specific performance score, which were obtained from a sub-analysis of the data collected for one study. The study is from 2010.
    - Six sites participated in the pilot study including 3 urban, 2 suburban, and 1 rural location in Florida, Wisconsin, Oregon, and Arizona. The number of cardiologists at each site ranged from 7 to 20 physicians. The number of SPECT MPI patients submitted from each site varied from 328 to 1,597 patients, but a total of 6,351 subjects with complete data were entered into the pilot database.
    - The developer provided results for four sites with results ranging from 0% to 1.2%. No specific information is provided about each of the site, i.e., size, number of studies, location, ownership, or the timeframe when the data were obtained.
    - There is not enough information to determine if the results provided correspond to the levels of analysis for which this measure is specified. The study only includes one of the four types of tests included in the measure.
- The developers present additional information from the literature demonstrating a range of rates of inappropriate cardiac stress testing in a range of clinical situations.

**Disparities**

- No information on disparities is provided.

***Questions for the Committee:***

- Does the developer provide enough data to show a gap in care that warrants a national performance measure?
- Does the data provided demonstrate a need for this measure?
- Since the developer did not provide any information on disparities, are you aware of evidence that disparities exist in this area of healthcare?

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

**RATIONALE:** The data provided for performance gap and disparities is minimal or insufficient. The data provided are from 2010, providing no information on current performance gaps. Performance scores on the measure as specified are required for maintenance of endorsement. Those scores are not provided.

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

1a.

- I agree that the evidence is moderate
- Evidence has face validity, but actual evidence is surprisingly weak
- Moderate evidence. Strong clinical rationale but little RCT data
- The evidence seems a little weak. Quick review of the guidelines show that the 2014 is the most recent, but hard to believe there hasn't been another trial.
- No new evidence was presented, though they state the evidence was Grade B Class III

1b.

- I don't believe that we know the current performance gap
- not provided
- No current data reported. Last data in 2010 from small number of practices
- Majority of data from a 2010 study. No gap analyzed from current data.
- There was limited data provided from 4 sites which showed scores ranging from 0.0%-1.2%.

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability: Specifications and Testing**

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

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

### Reliability

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

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

### Validity

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

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

**Composite measures only:**

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

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

**Evaluators:** NQF Staff

Scientific Acceptability: Preliminary Analysis Form

**Measure Number:** 0670

**Measure Title:** Cardiac stress imaging not meeting appropriate use criteria: Preoperative evaluation in low risk surgery patients

### 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: Unclear**

### Measure is:

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

### RELIABILITY: SPECIFICATIONS

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

  **Submission document:** "MIF_0670" 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.*

- **Briefly summarize any concerns about the measure specifications.**
  - No changes to the specifications were made from the previous submission in 2015.
  - It's unclear in the specifications what the data source is for each element or if there is more than one possible data source. Collection form included as an attachment does not appear to capture the CPT codes used to identify the numerator (in the measure submission form). If the codes are obtained from claims or EHR data, is there a maximum time frame between the test and the surgery (within 30 days after the cardiac test? 60 days? Etc.)?
  - Does the measure include all ages? No age range is included in the specifications.
  - The developer indicates Clinician: Group/Practice is a level of analysis for this measure. It's unclear what clinician would be held accountable. The denominator of number of tests

performed doesn't correspond to an ordering physician. Is it the performing physician? The attribution should be clear.

### RELIABILITY: TESTING

**Submission document:** "MIF_0670" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

- **Reliability testing level** ☐ **Measure score** ☒ **Data element** ☐ **Neither**
- **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☐ **Yes** ☒ **No**
    - Information supplied in testing attachment does not appear to correspond to data source or levels of analysis indicated.

- 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**

- **Assess the method(s) used for reliability testing**
    - The developer states reliability was tested at the data element level.
    - The study included to demonstrate reliability testing is a single-center study including 298 patients. It includes stress echocardiogram and SPECT MPI, but not the other cardiac tests included in the measure specifications.
    - The study included appears to focus on using appropriate use criteria to evaluate the appropriateness of a test whereas this measure attempts to identify tests used solely for pre-operative evaluation prior to low-risk surgery. It's unclear how precisely the appropriate use criteria in the study correspond to the measure specifications.
    - The inter-rater reliability provided is for the level of agreement in two nurses' appropriateness ratings for the cardiac testing. Appropriateness ratings are not a data element of this measure. The relationship between the appropriateness ratings and the measure specifications is unclear.

**Submission document:** Testing attachment, section 2a2.2

- **Assess the results of reliability testing**
    - There is not enough information provided to assess the reliability of this measure or its data elements.

**Submission document:** Testing attachment, section 2a2.3

- 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)

- 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)

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

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

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

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

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

- **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - There is not enough information provided to assess the reliability of this measure or its data elements. The information provided in the reliability section is not clearly related to the measure score or to the data elements in the measure. Testing does not appear to correspond to the levels of analysis (clinician: group/practice and facility) indicated for the measure.
    - In addition to concerns with the testing, staff identified concerns with the clarity of the specifications, particularly clinician attribution.

### VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

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

- This measure has no exclusions.

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

**Submission document:** Testing attachment, section 2b4.

- The developer's discussion of differences in performance focuses on inappropriate use. It is unclear if inappropriate use corresponds to results as calculated using this measure or if it is an application of the AUC.
- The developer noted that statistical tests have not been applied to demonstrate differences among the measured entities at the practice/hospital level.
- While the developer notes that there is variation in inappropriate use rates at the individual-practitioner level and that these rates vary by physician specialty, no method is highlighted to identify meaningful differences in performances.
- It isn't clear if the studies the developer is referencing are measuring the same results as this measure.

- **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

**Submission document:** Testing attachment, section 2b5.

- Not applicable

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

**Submission document:** Testing attachment, section 2b6.

- The developer reports "All subjects were classifiable according to the 2009 AUC and therefore no analysis for missing data was required." It is unclear which patients are being referenced

and the relationship between the data in the study and the data elements of this measure is unclear.

- **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

- Measure is not risk-adjusted.

## VALIDITY: TESTING

- **Validity testing level: ☐ Measure score ☐ Data element ☐ Both**
- **Method of establishing validity of the measure score:**

☐ **Face validity**

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

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

- **Assess the method(s) for establishing validity**

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

- The developer states their method of validity testing is the "relationship between appropriate use score and predictive value of SPECT MPI." This does not appear to align with face or empirical validity testing for measure 0670.

- **Assess the results(s) for establishing validity**

**Submission document: Testing attachment, section 2b2.3**

- The results presented do not provide information that can be used to assess the validity of this measure. There is not enough relevant information provided.

- **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)

- **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)

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

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

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

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

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

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

  The information provided in the validity section is not directly related to the measure score or to the measure's data elements. There is not enough information provided to assess the validity of the measure score or the data elements.

## ADDITIONAL RECOMMENDATIONS

- **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.**

***Questions for the Committee regarding reliability:***

- Is it clear from the provided specifications how this measure would be attributed to a clinician group or practice and which clinician group or practice would be held accountable?
- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The staff was not satisfied with the reliability testing for the measure. Does the Committee agree with the staff assessment?

***Questions for the Committee regarding validity:***

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The staff was not satisfied with the validity analyses for the measure. Does the Committee agree with the staff assessment?

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

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

2a1.

- The specifications are not clear
- Developer understands the attribution issues between ordering provider and performing lab, both should be reported and aggregated. Not so worried about time frame, use of a pre-op CPT without other codes to define numerator is enough
- Unclear measure specs: age? time between testing and surgery?
- Agree with NQF assessment of the reliability
- no answer

2a2.

- Yes. I don't think that the reliability testing is significant
- As constructed will be a conservative underestimation of inappropraite use.
- No. Inter-rater reliability measured at appropriateness level, not at data element level;
- Agree with NQF assessment of the reliability
- data element testing was done. Not clear if the testing was reported from the literature or from the sample of 4 sites. Testing was done between August 2007 and May 2010

2b1.

- Empiric validity testing is lacking
- Testing attachment not available at this time.
- No statistical testing so unclear what makes for meaningful differences in use, or how that is defined
- Agree with NQF assessment of the reliability
- unclear about the testing

2b4-7.

- The data have not been updated in the current application
- No data to review. Data on inappropriate use of various modalities would be enlightening. Why were plain treadmills not included? Are stress PET scans included?
- No clear statistical underpinning for performance scores
- Agree with NQF assessment of the reliability
- Not sure there are meaningful differences in quality.

2b2-3.

- see my comments at the end
- No data on socio-economic variables, no risk adjustment, no data to review
- No clear validity testing as applied to measure score
- Agree with NQF assessment of the reliability
- There are no exclusions or risk adjustment

## Criterion 3. Feasibility

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

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

- Data is generated by and used by healthcare personnel during the provision of care, (e.g., indications for testing) and coded by someone other than person obtaining the original information.
- Some data elements are in defined fields in electronic sources and some may require abstraction.

***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?
- What is the burden of data collection, i.e., chart abstraction and data entry to a registry?

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

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

- I agree that feasibility is moderate
- Limited data collection to a 60 day period (if adequate number of studies) to minimize burden
- No threats
- The data seem vague - either with how it is captured or obtained digitally. Not specific.
- seems feasible

## Criterion 4: Usability and Use

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

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

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

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

**Current uses of the measure**

**Publicly reported?** ☒ **Yes** ☐ **No**

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

**Accountability program details**

- The developer states the measures is used in the following programs:

- MIPS - CMS/pay for performance/national; The data collected at the lab level for this measure can be further segmented by physician to help them understand their appropriate use patterns; although small sample sizes can limit comparability for some providers
- FOCUS - ACC/lab accreditation, quality improvement and utilization management/national - 25,000 cases with concentrations in DE (100% for SPECT MPI) and Western PA (10% for SPECT MPI and stress echo for cardiologists) - additional 6,000 cases
- IAC - lab accreditation/national – this measure may be used in support of accreditation

**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**

- None.

**Additional Feedback:**

- None

***Questions for the Committee**:*

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

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

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

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

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

**Improvement results**

- The developer provides data from a cohort study of a 5% national sample of Medicare beneficiaries, annual rates of overall testing appeared to increase from 2000 to 2008 and then declined until 2016. Rates of low-value tests (preoperative stress testing and routine stress testing after coronary revascularization) appeared to have increased and then decreased.
- Trends of results for this specific measure were not provided.

**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** No unexpected findings provided.

**Potential harms** According to the developers, no unintended consequences have been identified for this measure.

**Additional Feedback:**

***Questions for the Committee**:*

- Are you aware of any unintended consequences for this measure?
- Do the benefits of the measure outweigh any potential unintended consequences?

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

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

4a1.

- The following info is highly relevant: The 2014 Protecting Access to Medicare Act included a provision requiring clinicians consult with appropriate use criteria (AUC) through a qualified clinical decision support mechanism (CDSM) when ordering advanced imaging services (i.e., SPECT/PET MPI, CT and MR) in order to receive payment approval from the Centers for Medicare and Medicaid Services (CMS). Following several delays, the implementation of this requirement started Jan. 1, 2020. ACC is working to educate and prepare providers for the implementation of potentially significant changes to practice workflows resulting from the AUC Program
- Not used enough, benefit of this measure requires feedback and trending.
- Used as part of MIPS
- In use in publicly reported programs.
- They say it is publically reported

4b1.

- see: https://www.acc.org/tools-and-practice-support/quality-programs/imaging-in-focus/auc-mandate-information
- no harms, usability potential is high, not yet demonstrated
- No harms of note
- in 4b2.1 the steward discusses clinical judgement to determine list of surgeries. Would like to see more specificity if possible.
- unclear about harms

## Criterion 5: Related and Competing Measures

**Related or competing measures**

- 0669: Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac, Low Risk Surgery
- 0671: Cardiac stress imaging not meeting appropriate use criteria: Routine testing after percutaneous coronary intervention (PCI)
- 0672: Cardiac stress imaging not meeting appropriate use criteria: Testing in asymptomatic, low risk patients

**Harmonization**

- The developer stated that harmonization with 0669 hasn't happened due to different populations and data sources used.

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

- 0669 . According to the worksheet, 0670 measure provides an additional level of analysis that applies not only to hospitals but also outpatient physician clinics. The data source also provides a richer source of clinical information to distinguish between testing ordered for preoperative assessment and other cardiovascular causes co-existing at the same time
- 669 may compete
- 0669: Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac, Low Risk Surgery. Per developer different populations/data sources
- Not harmonized with other measures.

## Public and Member Comments

**Comments and Member Support/Non-Support Submitted as of: January 21, 2020**

- **No NQF members have submitted a support/non-support choice as of this date.**

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

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

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

NQF_evidence_attachment_Sep2017_-_670-637099381169840536.docx

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

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

No.

### 1a. Evidence (subcriterion 1a)

**Measure Number** (*if previously endorsed*)**:** 670

**Measure Title**: Cardiac stress imaging not meeting appropriate use criteria: Preoperative evaluation in low risk surgery patients

**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**: 11/7/2019

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

Outcome

☐ Outcome: Resource Use and Avoidance of Negative Clinical Benefit Risk Ratio for Patient Population

☐ 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: Resource Use and Avoidance of Negative Clinical Benefit Risk Ratio for Patient Population

☒ Appropriate use measure: Cardiac stress imaging not meeting appropriate use criteria: Preoperative evaluation in low risk surgery patients

☐ 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.

Measurement of appropriate use summarizes the financial value/resources use and avoidance of a negative clinical benefit risk ratio across a patient population in which a procedure is used. Various factors influence the ability of a procedure to contribute to the diagnosis and treatment of a patient, including the clinical factors summarized by appropriate use measures. These clinical factors combined with physician and patient decision making determine the probability that a procedure will have the intended impact on health outcomes of the patient

**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.**

**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*)

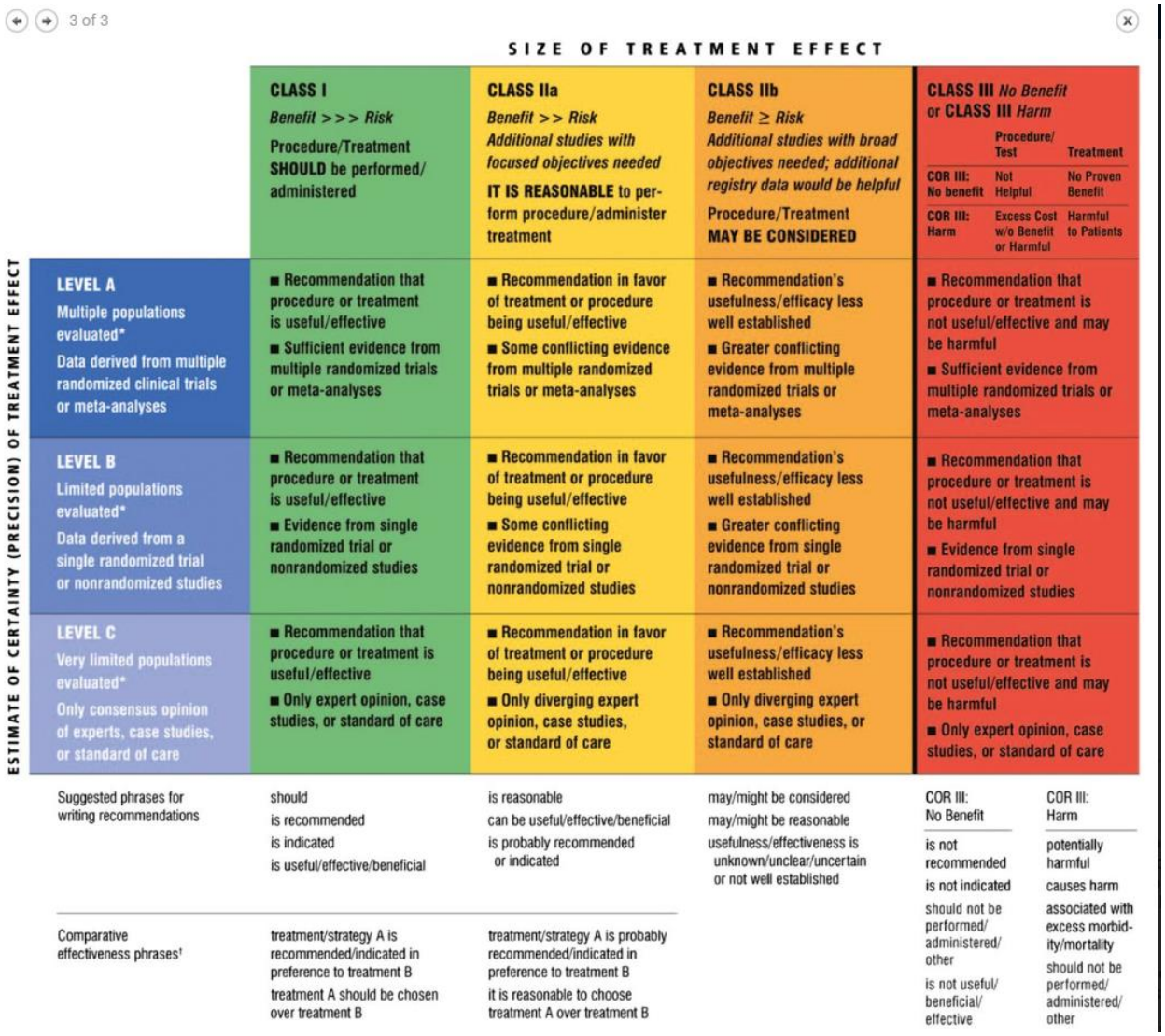
☒ Other

| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | **2014 ACC/AHA Guideline on Perioperative Cardiovascular Evaluation and Management of Patients Undergoing Noncardiac Surgery**<br>**A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines**<br><br>**Lee A. Fleisher, Kirsten E. Fleischmann, Andrew D. Auerbach, Susan A. Barnason, Joshua A. Beckman, Biykem Bozkurt, Victor G. Davila-Roman, Marie D. Gerhard-Herman, Thomas A. Holly, Garvan C. Kane, Joseph E. Marine, M. Timothy Nelson, Crystal C. Spencer, Annemarie Thompson, Henry H. Ting, Barry F. Uretsky and Duminda N. Wijeysundera**<br><br>**December 2014**<br><br>**J Am Coll Cardiol. 2014 Dec, 64 (22) e77-e137**<br><br>**http://www.onlinejacc.org/content/64/22/e77** |

| | |
|---|---|
| | |

| | |
|---|---|
| **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.** | **e97**<br>**CLASS III: NO BENEFIT 1. Routine screening with noninvasive stress testing is not useful for patients undergoing low-risk noncardiac surgery (165,166). (Level of Evidence: B)**<br><br>**165.**<br>**Sgura F.A., Kopecky S.L., Grill J.P., et al; Supine exercise capacity identifies patients at low risk for perioperative cardiovascular events and predicts long-term survival. Am J Med. 2000;108:334-336.**<br><br>**166.**<br>**Mangano D.T., London M.J., Tubau J.F., et al; Dipyridamole thallium-201 scintigraphy as a preoperative screening test: a reexamination of its predictive potential. Study of Perioperative Ischemia Research Group. Circulation. 1991;84:493-502** |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | Level of Evidence: B<br>Data derived from a single randomized trial, or non-randomized studies |
| Provide all other grades and definitions from the evidence grading system | See below* |
| Grade assigned to the **recommendation** with definition of the grade | CLASS III: NO BENEFIT<br>Conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective, and in some cases may be harmful. |
| Provide all other grades and definitions from the recommendation grading system | See below* |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | Quantity: 25 single center observational studies<br><br>Quality: Only a few of the studies addressed the surgical population focused on in this measure. As such, the event rates are much higher among the population in the majority of the studies given they examined higher risk surgeries. Studies examining non-vascular surgeries had much lower events rates but still did not limit their populations to low risk surgery. |

| | |
|---|---|
| **Estimates of benefit and consistency across studies** | **This measure looks at the absence of potential benefit in a specific population which is derivative of the studies examined but not a direct end point of the studies reviewed. The general association of lower mortality for lower risk surgeries and higher mortality for higher risk surgeries was consistent across the studies.** |
| What harms were identified? | The studies did not examine harm of providing the imaging procedure. |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | The below** studies were reviewed by the guideline writing committee and are provided for informational purposes. They are not new since the guideline. |



3 of 3

SIZE OF TREATMENT EFFECT

ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT

| | CLASS I<br>*Benefit >>> Risk*<br>Procedure/Treatment **SHOULD** be performed/ administered | CLASS IIa<br>*Benefit >> Risk*<br>*Additional studies with focused objectives needed*<br>**IT IS REASONABLE** to perform procedure/administer treatment | CLASS IIb<br>*Benefit ≥ Risk*<br>*Additional studies with broad objectives needed; additional registry data would be helpful*<br>Procedure/Treatment **MAY BE CONSIDERED** | CLASS III *No Benefit* or CLASS III *Harm*<br>Procedure/ Test / Treatment<br>COR III: No benefit / Not Helpful / No Proven Benefit<br>COR III: Harm / Excess Cost w/o Benefit or Harmful / Harmful to Patients |
|---|---|---|---|---|
| **LEVEL A**<br>Multiple populations evaluated*<br>Data derived from multiple randomized clinical trials or meta-analyses | ■ Recommendation that procedure or treatment is useful/effective<br>■ Sufficient evidence from multiple randomized trials or meta-analyses | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Some conflicting evidence from multiple randomized trials or meta-analyses | ■ Recommendation's usefulness/efficacy less well established<br>■ Greater conflicting evidence from multiple randomized trials or meta-analyses | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Sufficient evidence from multiple randomized trials or meta-analyses |
| **LEVEL B**<br>Limited populations evaluated*<br>Data derived from a single randomized trial or nonrandomized studies | ■ Recommendation that procedure or treatment is useful/effective<br>■ Evidence from single randomized trial or nonrandomized studies | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Some conflicting evidence from single randomized trial or nonrandomized studies | ■ Recommendation's usefulness/efficacy less well established<br>■ Greater conflicting evidence from single randomized trial or nonrandomized studies | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Evidence from single randomized trial or nonrandomized studies |
| **LEVEL C**<br>Very limited populations evaluated*<br>Only consensus opinion of experts, case studies, or standard of care | ■ Recommendation that procedure or treatment is useful/effective<br>■ Only expert opinion, case studies, or standard of care | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Only diverging expert opinion, case studies, or standard of care | ■ Recommendation's usefulness/efficacy less well established<br>■ Only diverging expert opinion, case studies, or standard of care | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Only expert opinion, case studies, or standard of care |
| Suggested phrases for writing recommendations | should<br>is recommended<br>is indicated<br>is useful/effective/beneficial | is reasonable<br>can be useful/effective/beneficial<br>is probably recommended or indicated | may/might be considered<br>may/might be reasonable<br>usefulness/effectiveness is unknown/unclear/uncertain or not well established | COR III: No Benefit<br>is not recommended<br>is not indicated<br>should not be performed/ administered/ other<br>is not useful/ beneficial/ effective<br>COR III: Harm<br>potentially harmful<br>causes harm<br>associated with excess morbidity/mortality<br>should not be performed/ administered/ other |
| Comparative effectiveness phrases† | treatment/strategy A is recommended/indicated in preference to treatment B<br>treatment A should be chosen over treatment B | treatment/strategy A is probably recommended/indicated in preference to treatment B<br>it is reasonable to choose treatment A over treatment B | | |

***** grades and definitions from the evidence grading system

**STUDIES

**Data Supplement 14. Radionuclide MPI (Section 5.5.2)**

| Study Name, Author, Year | Aim of Study | Study Type | Study Size (N) | Patient Population | | Ischemia | |
|---|---|---|---|---|---|---|---|
| | | | | *Inclusion Criteria* | *Exclusion Criteria* | | *Primary Endpoint (Efficacy) and Results* |
| Eagle KA, et al.,1989 (77) 8653858 | Periop risk assessment by MPI | Single center, retrospective | 200 | Vascular surgery | N/A | 41% | Periop events: PPV: 16%; NPV: 98% |
| Younis LT, et al., 1990 (78) 2353615 | Periop risk assessment by MPI | Single center, retrospective | 111 | Peripheral vascular disease | N/A | 36% | Periop events: PPV: 15%; NPV: 100% |
| Hendel RC, et al., 1992 (79) 1442573 | Periop risk assessment by MPI | Single center, retrospective | 327 | N/A | N/A | 51% | Periop events: PPV: 14%; NPV: 99% |
| Lette J, et al., 1992 (80) 1598869 | Periop risk assessment by MPI | Single center, retrospective | 355 | N/A | N/A | 45% | Periop events: PPV: 17%; NPV: 99% |
| Brown KA, et al., 1993 (81) 8425993 | Periop risk assessment by MPI | Single center, retrospective | 231 | N/A | N/A | 33% | Periop events: PPV: 13%; NPV: 99% |
| Bry JD, et al., 1994 (82) 8301724 | Periop risk assessment by MPI | Single center, retrospective | 237 | N/A | N/A | 46% | Periop events: PPV: 11%; NPV: 100% |
| Marshall ES, et al., 1995 (83) 7572662 | Periop risk assessment by MPI | Single center, retrospective | 117 | N/A | N/A | 47% | Periop events: PPV: 16%; NPV: 97% |
| Stratman HG, et al., 1996 (84) 8615311 | Periop risk assessment by MPI | Single center, retrospective | 229 | Nonvascvular surgery | N/A | 29% | Periop events: PPV: 6%; NPV: 99% |
| Cohen MC, et al., 2003 (85) 14569239 | Periop risk assessment by MPI | Single center, retrospective | 153 | N/A | N/A | 31% | Periop events: PPV: 4%; NPV: 100% |
| Harafuji K, et al., 2005 (86) 15849442 | Periop risk assessment by MPI | Single center, retrospective | 302 | N/A | N/A | 30% | Periop events: PPV: 2%; NPV: 100% |
| Beattie WS, et al., 2006 (76) 16368798 | Compare SE vs. MPI in preop evaluation prior to noncardiac surgery | Meta-analysis of 68 studies | 10,049 | Preop noncardiac surgery | N/A | N/A | Outcomes: MI and/or death |

**Data Supplement 15. Dobutamine Stress Echocardiography (Section 5.5.3)**

| Study Name, Author, Year | Aim of Study | Study Type | Study Size (N) | Patient Population | Events (MI/death) | Ischemia, % | Endpoints | | P Values, OR: HR: RR & 95% CI: | Study Limitations & Adverse Events |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | *Inclusion Criteria* | | | *Primary Endpoint (Efficacy) and Results* | *Secondary Endpoint and Results* | | |
| Lane RT, et al.,1991 (87) 1927965 | Periop risk assessment by DSE | Single center, retrospective | 38 | Vascular and general surgery | 8% | 50% | PPV 16%, NPV 100% | N/A | N/A | N/A |
| Lalka SG. et al., 1992 (88) 1578539 | Periop risk assessment by DSE | Single center, retrospective | 60 | Abdominal aortic surgery | 15% | 50% | PPV 23%, NPV 93% | N/A | Event rate 29% vs. 4.6%, p=0.025 | N/A |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Eichelberger JP, et al., 1993 (89) 8362778 | Periop risk assessment by DSE | Single center, prospective | 75 | Major vascular surgery | 3% | 36% | PPV 7%, NPV 100% | N/A | N/A | N/A |
| Langan EM, et al., 1993 (90) 8264046 | Periop risk assessment by DSE | Single center, retrospective | 74 | Aortic surgery | 4% | 24% | PPV 17%, NPV 100% | N/A | N/A | Surgery deferred in 4 highly positive DSE who proceeded with CABG |
| Davila-Roman V, et al., 1993 (91) 8450165 | Periop risk assessment by DSE | Single center, prospective | 88 | Aortic and LE PVD surgery | 2% | 23% | PPV 10%, NPV 100% | Abnormal DSE associated with increased long-term event rate also (15% vs. 3%; p=0.038) | N/A | N/A |
| Shafritz R, et al., 1997 (92) 9293826 | Periop risk assessment by DSE, comparison to historical cohort without preop DSE | Single center, retrospective | 42 | Aortic surgery | 2% | 0% | NPV 100% | No difference in overall mortality (2.3% vs. 4.4%) or cardiac mortality (0% vs. 2.9%) in those who had preop DSE testing vs. those who did not | N/A | N/A |
| Bossone, 1999 (93) 10469973 | Periop risk assessment by DSE | Single center, prospective | 46 | Lung-volume reduction surgery | 2% | 9% | PPV 25%, NPV 100% | N/A | N/A | N/A |
| Ballal RS, et al., 1999 (94) 10047628 | Periop risk assessment by DSE | Single center, prospective | 233 | Major vascular surgery | 3% | 17% | PPV 0%, NPV 96% | N/A | N/A | Surgery deferred in 8 highly positive DSE who proceeded with PCI |
| Das MK, et al., 2000 (95) 10807472 | Periop risk assessment by DSE | Single center, prospective | 530 | Nonvascular surgery | 6% | 40% | PPV 15%, NPV 100% | High risk study (defined as ischemia before 60% of age-predicted heart rate threshold) associated event rate of 43%. Incremental risk prediction over clinical characteristics | N/A | N/A |
| Morgan PB, et al., 2002 (96) 12198027 | Periop risk assessment by DSE | Single center, retrospective | 78 | Vascular and general surgery | 0% | 5% | PPV 0, NPV 100% | N/A | N/A | All 4 pts with ischemia underwent precp coronary angiography +/- PCI. |
| Torres MR et al., 2002 (97) 12127610 | Periop risk assessment by DSE | Single center, prospective | 105 | Predominantly vascular surgery | 10% | 47% | PPV 18%, NPV 98% | N/A | N/A | Beta-blocker therapy given on basis of DSE, 4 pts had surgery deferred for PCI/CABG |
| Lakib SB, et al., 2004 (98) 15234412 | Periop risk assessment by DSE, comparison of maximal vs. submaximal achieved peak heart rate | Single center, prospective | 429 | 1/3 vascular surgery | 2% | 7% | PPV 9%, NPV 98% | High NPV even when peak heart rate not achieved | N/A | N/A |
| Raux M, et al., 2006 (99) | Periop risk assessment by a | Single center, retrospective | 143 | Abdominal aortic surgery | N/A | N/A | NPV 93% events predominantly were | N/A | N/A | All with abnormal DSE underwent coronary |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 16973646 | negative DSE and incidence of elevated troponin | | | | | | nonclinical elevated troponin measures | | | angiogram +/- PCI prior to surgery |
| Umphrey LG, et al., 2008 (100) 18508373 | Periop risk assessment by DSE | Single center, retrospective | 157 | Orthotropic liver transplantation | 3.80% | 0% | NPV | Inability during DSE to achieve >80% of targeted heart rate associated with increased cardiac events (22% vs. 6%; p=0.01) | N/A | N/A |
| Lerakis S, et al., 2007 (101) 18219774 | Periop risk assessment by DSE | Single center, retrospective | 539 | Bariatric surgery | 0.05% (all noncardiac death) | 1.20% | N/A | N/A | N/A | All with abnormal DSE underwent coronary angiogram +/- PCI prior to surgery |
| Nguyen P, et al., 2013 23974907 | Periop risk assessment by DSE | Pooled analysis of 7 studies | 580 | Orthotropic liver transplantation | N/A | N/A | PPV 37%, NPV 75% | N/A | N/A | N/A |

## 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.*

Measurement of appropriate use summarizes the financial value/resources use and avoidance of a negative clinical benefit risk ratio across a patient population in which a procedure is used. Various factors influence the ability of a procedure to contribute to the diagnosis and treatment of a patient, including the clinical factors summarized by appropriate use measures. These clinical factors combined with physician and patient decision making determine the probability that a procedure will have the intended impact on health outcomes of the patient.

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Circulation. 2013 Oct 8;128(15):1634-43. doi: 10.1161/CIRCULATIONAHA.113.002744. Epub 2013 Sep 10.

Patients in the appropriate/uncertain group experienced significantly higher overall rates of death (HR, 2.9; 95% CI, 1.05–8.0; P=0.04), the composite of death or MI (HR, 1.04; 95% CI, 1.01–1.07; P=0.03), and the composite of cardiac death or MI (HR=5.7; 95% CI, 1.3–25.6; P=0.02) after adjustment for clinical covariates. Among patients in the appropriate/uncertain group, abnormal MPI continued to predict a multifold increase in the risk of death, cardiac death, composite of death or MI, and composite of cardiac death or MI (Figure 4).

However, in the inappropriate group, there were no statistically significant differences in MACE rates between subjects with abnormal versus normal MPI (Figure 4). Furthermore, using Cox regression models, no interaction was identified between the study group and MPI finding in predicting death, the composite of death or MI, or the composite of cardiac death or MI (P=0.91, 0.70, and 0.43, respectively).

A Cox regression model demonstrated that inappropriate MPI use was a negative predictor of all-cause mortality (HR, 0.26; 95% CI, 0.10–0.67; P=0.005) after adjustment for myocardial perfusion finding (normal versus abnormal; HR, 2.5; 95% CI, 1.1–5.9; P=0.04) and depressed LVEF (<50%; HR, 3.7; 95% CI, 1.5–9.3; P=0.006); undergoing early coronary revascularization was not predictive of mortality (P=0.98). Similarly, in separate models, we demonstrated that inappropriate use was an independent negative predictor of the secondary end points of death or MI (HR, 0.31; 95% CI, 0.14–0.70; P=0.005) and cardiac death or MI (HR, 0.16; 95% CI, 0.04–0.71; P=0.02) after adjustment for depressed LVEF, myocardial perfusion findings, and early revascularization. In these models, MPI and depressed LVEF independently predicted the composite end points of death or MI and cardiac death or MI, whereas undergoing early coronary revascularization after MPI was not predictive of these end points (P≥0.97). Finally, in forward stepwise Cox regression models, appropriate use was shown to have incremental prognostic value to perfusion imaging and depressed LVEF in predicting MACE; undergoing early revascularization (<60 days) did not provide significant additional predictive value

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

Each of the documents below covers a clinical imaging procedure and was developed using the AUC methodology above.

The Appropriate Use Criteria have been published and updated on a regular basis by the American College of Cardiology in partnership with other societies and stakeholders. The evidence underlying the AUC appear in guidelines and systematic reviews contained in the appendix materials for these documents. The clinical indications and expert opinion used have been widely studied for their applicability to imaging rationale as well as outcomes.

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

A rigorous and validated process involving multiple societies and other stakeholders was used to develop the Appropriate Use Criteria (AUC). The AUC have been validated by various studies, including the ones cited earlier in this application. They are not merely expert panels but purposefully balanced committees undergoing a rigorous consensus process beyond even those used by guideline panels for decision making. A RAND modified Delphi process is used to determine the AUC rating that combines expert opinion with available evidence and specific patient information. The methods for this review have been published and are available at: http://www.onlinejacc.org/content/71/8/935?_ga=2.169985062.746725178.1574208699-1575853885.1561572054 and https://www.acc.org/guidelines#tab4. Few studies are conducted to demonstrate a lack of benefit and thus, clinical risk and expert opinion is required to develop the AUC. Guidelines on the topic and references supporting recommendations related to the AUC clinical indications were identified. Additional literature searches were conducted to complete the available evidence published since the last guideline update. Specific evidence grades are not assigned by AUC, but generally diagnostic imaging evidence is based on observational studies, including well known risk models such as Framingham and Diamond and Forrester. In addition, a RAND modified Delphi process is used to determine the AUC rating that combines expert opinion with available evidence and specific patient information.

Few studies are conducted to demonstrate a lack of benefit and thus, clinical risk and expert opinion is required to develop the AUC

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

Original

Douglas PS, Khandheria B, Stainback RF, ACCF/ASE/ACEP/AHA/ASNC/SCAI/SCCT/SCMR2008 appropriateness criteria for stress echocardiography. J Am Coll Cardiol. 2008;51:1127–47.

Hendel RH, Berman DS, Di Carli MF, et al. ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 Appropriate Use Criteria for Cardiac Radionuclide Imaging. J Am Coll Cardiol. 2009;53:2201–29.

Hendel RC, Patel MR, Kramer CM, Poon M. ACCF/ACR/SCCT/SCMR/ASNC/NASCI/SCAI/SIR 2006 appropriateness criteria for cardiac computed tomography and cardiac magnetic resonance imaging. J Am Coll Cardiol 2006;48:1475–97.

Updated

Wolk MJ, Bailey SR, Doherty JU et al. ACCF/AHA/ASE/ASNC/HFSA/HRS/SCAI/SCCT/SCMR/STS 2013 multimodality appropriate use criteria for the detection and risk assessment of stable ischemic heart disease. J Am Coll Cardiol 2014;63:XXX–XX.

The Appropriate Use Criteria have been published and updated on a regular basis by the American College of Cardiology in partnership with other societies and stakeholders. The evidence underlying the AUC appear in guidelines and systematic reviews contained in the appendix materials for these documents. The clinical indications and expert opinion used have been widely studied for their applicability to imaging rationale as well as outcomes.

## 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.*

Appropriate use criteria define "when to do" and "how often to do" a given procedure in the context of scientific evidence, the health care environment, the patient's profile and a physician's judgment. While practice guidelines provide a foundation for summarizing evidence-based cardiovascular care or for providing expert consensus opinions, in many areas, marked variability remains in the use of cardiovascular procedures, raising questions about over-use and under-use. Appropriate use criteria provide practical tools to measure this variability and to look at utilization patterns. The criteria are designed to examine the use of diagnostic and therapeutic procedures to support efficient use of medical resources, while also providing patients with quality, appropriate care.

A measure that reports rates of inappropriate imaging within practices would contain information regarding both cost and quality, because an inappropriate test results in both higher costs and poorer-quality care.

Conversely, a reduction in this rate would simultaneously improve quality and decrease cost. Improvements in this metric should lead to consistent application of AUC and improve the efficiency of the system.

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

Hendel RC, Cerqueira M, Douglas PS. A multicenter assessment of the use of single-photon emission computed tomography myocardial perfusion imaging with appropriateness criteria. J Am Coll Cardiol. 2010 Jan 12;55(2):156-62.

These site specific performance scores were provided by a sub-analysis of the data collected for the above study. While the rates are fairly low in these sites, the additional studies in 1b.3 demonstrate a range of performance on this measure that is generally higher than these rates.

Six sites participated in this pilot study; 3 urban, 2 suburban, and 1 rural location. Practices were located in Florida, Wisconsin, Oregon, and Arizona, and the number of cardiologists at each site ranged from 7 to 20 physicians. The number of SPECT MPI patients submitted from each site varied from 328 to 1,597 patients. A total of 6,351 subjects with complete data were entered into the pilot database.

All Sites - 0.5%

Site 1 - 1.1%

Site 2 - 1.2%

Site 3 - 1.1%

Site 4 - 0.0%

**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.**

Hendel RC, Cerqueira M, Douglas PS. A multicenter assessment of the use of single-photon emission computed tomography myocardial perfusion imaging with appropriateness criteria. J Am Coll Cardiol. 2010 Jan 12;55(2):156-62.

See above.

Fonseca R, Negishi K, Otahal P, et al. Temporal Changes in Appropriateness of Cardiac Imaging. J Am Coll Cardiol. 2015 Mar 3;65(8):763-73.

Mehta R, Agarwal S, Chandra S, Ward RP, Williams KA: Evaluation of the American College of Cardiology Foundation/American Society of Nuclear Cardiology appropriateness criteria for SPECT myocardial perfusion imaging. J Nucl Cardiol. 2008;5:337–44.

There were 1,623 patients (mean age 61 years ± 11, 61% males). Most common indications for SPECT were evaluation of ischemic equivalent for coronary artery disease (CAD), risk assessment post-revascularization, and preoperative evaluation for non-cardiac surgery. 10% of referrals were classified as inappropriate, 5% uncertain, and 3% unclassified. Appropriate referrals had a higher proportion of abnormal SPECT results than inappropriate referrals (40% vs 27%, OR 2.08, 95% CI 1.56-2.77, P < .001).

Ward RP, Al-Mallah MH, Grossman GB, Hansen CL, Hendel RC, Kerwin TC, McCallister BD Jr., Mehta R, Dm Polk, Tilkemeier PL,Vashist A, Williams KA, Wolinsky DG, Ficaro EP: American Society of Nuclear Cardiology: American Society of Nuclear Cardiology review of the ACCF/ASNC appropriateness criteria for single-photon emission computed tomography myocardial perfusion imaging

SPECT MPI). J Nucl Cardiol. 2007;14:e26–38.

Gibbons RJ, Miller TD, Hodge D, Urban L, Araoz PA, Pellikka P, McCully RB: Application of appropriateness criteria to stress single photon emission computed tomography sestamibi studies and stress echocardiograms in an academic medical center. J Am Coll Cardiol. 2008;51:1283–9.

The purpose of this study was to apply published appropriateness criteria for single-photon emission computed tomography (SPECT) myocardial perfusion imaging (MPI) in a single academic medical center.

The study retrospectively examined 284 patients who underwent stress SPECT MPI and 298 patients who underwent stress echocardiography before publication of the criteria. 10% of studies were for evaluation prior to low risk surgery.

Sheffield KM, McAdams PS, Benarroch-Gampel J, Goodwin JS, Boyd CA, Zhang D, Riall TS. Overuse of preoperative cardiac stress testing in medicare patients undergoing elective noncardiac surgery. Ann Surg. 2013 Jan;257(1):73-80.

In a 5% sample of Medicare claims data, 2803 patients underwent preoperative stress testing without any indications. When these results were applied to the entire Medicare population, we estimated that there are over 56,000 patients who underwent unnecessary preoperative stress testing. The rate of testing in patients without cardiac indications has increased significantly over time.

Carryer DJ, Hodge DO, Miller TD, Askew JW, Gibbons RJ. Application of appropriateness criteria to stress single photon emission computed tomography sestamibi studies: a comparison of the 2009 revised appropriateness criteria to the 2005 original criteria. Am Heart J. 2010 Aug;160(2):244-9..

An equal percentage of inappropriate studies (10.3%) were due to pre-op evaluation for low to intermediate risk non-cardiac surgery (indications 40, 41, and 42) and asymptomatic patients, less than 2 years after PCI (indication no. 59).

MedPAC Report to the Congress: Medicare and the Health Care Delivery System. Chapter 3. Measuring quality of care in Medicare. June 2014

Rates ranged from 4.7% to 5.3% of imaging performed for evaluation prior to low risk surgery.

**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.*

None

**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**

this doesn´t seem to exist. seems like an opportunity.

## 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):*

Cardiovascular, Cardiovascular : Coronary Artery Disease

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

Safety : Overuse

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

Elderly

**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.)*

None at this time except NQF specifications page

**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:** Imaging-Efficiency-Measures-Micro-specifications_Measure_Maintenance-635231526161153276.doc

**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.

no changes to specifications.

**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).*

Number of stress SPECT MPI, stress echo, CCTA, or CMR performed in patients undergoing low risk surgery as a part of the preoperative evaluation

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

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

Patients qualify this measure if:

-an upcoming surgery is the recorded reason for the imaging test AND

-no other reason is recorded for the imaging

AND

Surgery risk is low

The following will be used to determine whether the risk of the surgery recorded is low:

Surgical Risk Categories

• Low-Risk Surgery– cardiac death or MI less than 1% including endoscopic procedures, superficial procedures, cataract surgery, breast surgery.

Surgeries meeting this definition to be included in the measure are listed by

CPT 4 Codes below. While additional surgeries may fit the low risk definition, only those surgeries listed below will be considered in determining inclusion in the numerator for this measure.

Surgery/Integumentary System: Breast

19100 Biopsy of breast

19101 Biopsy of breast

19102 Bx breast percut w/image

19103 Bx breast percut w/device

Surgery/Respiratory System: Accessory Sinuses

31231 Nasal endoscopy, dx

31233 Nasal/sinus endoscopy, dx

31235 Nasal/sinus endoscopy, dx

31237 Nasal/sinus endoscopy, surg

31238 Nasal/sinus endoscopy, surg

31239 Nasal/sinus endoscopy, surg

31240 Nasal/sinus endoscopy, surg

31267 Endoscopy, maxillary sinus

31276 Sinus surgical endoscopy

31299 Sinus surgery procedure

Surgery/Respiratory System: Larynx

31505 Diagnostic laryngoscopy

31510 Laryngoscopy with biopsy

31511 Remove foreign body, larynx

31513 Injection into vocal cord

31515 Laryngoscopy for aspiration

31520 Diagnostic laryngoscopy

31525 Diagnostic laryngoscopy

31526 Diagnostic laryngoscopy

31527 Laryngoscopy for treatment

31528 Laryngoscopy and dilatation
31529 Laryngoscopy and dilatation
31530 Operative laryngoscopy
31531 Operative laryngoscopy
31535 Operative laryngoscopy
31536 Operative laryngoscopy
31540 Operative laryngoscopy
31541 Operative laryngoscopy
31560 Operative laryngoscopy
31561 Operative laryngoscopy
31570 Laryngoscopy with injection
31571 Laryngoscopy with injection
31575 Diagnostic laryngoscopy
31576 Laryngoscopy with biopsy
31577 Remove foreign body, larynx
31578 Removal of larynx lesion
31579 Diagnostic laryngoscopy
Surgery/Respiratory System: Trachea and Bronchi
31615 Visualization of windpipe
31620 Endobronchial us add-on
31622 Diagnostic bronchoscopy
31623 Dx bronchoscope/brush
31624 Dx bronchoscope/lavage
31625 Bronchoscopy with biopsy
31628 Bronchoscopy with biopsy
31629 Bronchoscopy with biopsy
31632 Bronchoscopy/lung bx, add'l
31633 Bronchoscopy/needle bx add'l
31645 Bronchoscopy, clear airways
31646 Bronchoscopy, reclear airways
Surgery/Respiratory System: Lungs and Pleura
33508 Endoscopic vein harvest
37500 Endoscopy ligate perf veins
37501 Vascular endoscopy procedure
39400 Visualization of chest
Surgery/Digestive System: Esophagus
43200 Esophagus endoscopy
43201 Esophagus endoscopy, w/submucous injection
43202 Esophagus endoscopy, biopsy

43204 Esophagus endoscopy & inject
43205 Esophagus endoscopy/ligation
43215 Esophagus endoscopy
43216 Esophagus endoscopy/lesion
43217 Esophagus endoscopy
43219 Esophagus endoscopy
43220 Esophagus endoscopy, dilation
43226 Esophagus endoscopy, dilation
43227 Esophagus endoscopy, repair
43228 Esophagus endoscopy, ablation
43231 Esoph endoscopy w/us exam
43232 Esoph endoscopy w/us fn bx
43234 Upper GI endoscopy, exam
43235 Upper GI endoscopy, diagnosis
43236 Upper GI scope w/submuc inj
43237 Endoscopic us exam, esoph
43238 Upper GI endoscopy w/us fn bx
43239 Upper GI endoscopy, biopsy
43241 Upper GI endoscopy with tube
43242 Upper GI endoscopy w/us fn bx
43243 Upper GI endoscopy & inject.
43244 Upper GI endoscopy/ligation
43246 Place gastrostomy tube
43247 Operative upper GI endoscopy
43248 Upper GI endoscopy/guidewire
43249 Esophagus endoscopy, dilation
43260 Endoscopy, bile duct/pancreas
43261 Endoscopy, bile duct/pancreas
43262 Endoscopy, bile duct/pancreas
43263 Endoscopy, bile duct/pancreas
43264 Endoscopy, bile duct/pancreas
43265 Endoscopy, bile duct/pancreas
43267 Endoscopy, bile duct/pancreas
43268 Endoscopy, bile duct/pancreas
43269 Endoscopy, bile duct/pancreas
43271 Endoscopy, bile duct/pancreas
43272 Endoscopy, bile duct/pancreas
Surgery/Digestive System: Intestines (Except Rectum)
44360 Small bowel endoscopy

44361 Small bowel endoscopy, biopsy
44363 Small bowel endoscopy
44383 Ileoscopy w/stent
44385 Endoscopy of bowel pouch
44386 Endoscopy, bowel pouch, biopsy
44388 Colon endoscopy
44389 Colonoscopy with biopsy
44390 Colonoscopy for foreign body
44391 Colonoscopy for bleeding
44392 Colonoscopy & polypectomy
44393 Colonoscopy, lesion removal
44397 Colonoscopy w stent
Surgery/Digestive System: Rectum
45300 Proctosigmoidoscopy
45303 Proctosigmoidoscopy
45305 Proctosigmoidoscopy; biopsy
45307 Proctosigmoidoscopy
45308 Proctosigmoidoscopy
45309 Proctosigmoidoscopy
45315 Proctosigmoidoscopy
45317 Proctosigmoidoscopy
45320 Proctosigmoidoscopy
45321 Proctosigmoidoscopy
45327 Proctosigmoidoscopy w/stent
45330 Sigmoidoscopy, diagnostic
45331 Sigmoidoscopy and biopsy
45332 Sigmoidoscopy
45333 Sigmoidoscopy & polypectomy
45334 Sigmoidoscopy for bleeding
45335 Sigmoidoscope w/submuc inj
45337 Sigmoidoscopy, decompression
45338 Sigmoidoscopy
45339 Sigmoidoscopy
45340 Sig w/balloon dilation
45341 Sigmoidoscopy w/ultrasound
45342 Sigmoidoscopy w/us guide bx
45345 Sigmoidoscopy w/stent
45378 Diagnostic colonoscopy
45379 Colonoscopy

45380 Colonoscopy and biopsy
45381 Colonoscope, submucous inj
45382 Colonoscopy, control bleeding
45383 Colonoscopy, lesion removal
45384 Colonoscopy
45385 Colonoscopy, lesion removal
45387 Colonoscopy w/stent
45391 Colonoscopy w/endoscope us
45392 Colonoscopy w/endoscopic fnb
Surgery/Digestive System: Anus
46600 Diagnostic anoscopy
46604 Anoscopy and dilation
46606 Anoscopy and biopsy
46608 Anoscopy; remove foreign body
46610 Anoscopy; remove lesion
46612 Anoscopy; remove lesions
46614 Anoscopy; control bleeding
Surgery/Digestive System: Biliary Tract
47561 Laparo w/cholangio/biopsy
Surgery/Digestive System: Abdomen, Peritoneum and Omentum
49322 – Laparoscopy, aspiration
Surgery/Urinary System: Kidney
50551 Kidney endoscopy
50553 Kidney endoscopy
50555 Kidney endoscopy & biopsy
50557 Kidney endoscopy & treatment
50559 Renal endoscopy; radiotracer
50561 Kidney endoscopy & treatment
• Surgery/Urinary System: Ureter
50951 Endoscopy of ureter
50953 Endoscopy of ureter
50955 Ureter endoscopy & biopsy
50970 Ureter endoscopy
50972 Ureter endoscopy & catheter
50974 Ureter endoscopy & biopsy
50976 Ureter endoscopy & treatment
50978 Ureter endoscopy & tracer
50980 Ureter endoscopy & treatment
Surgery/Urinary System: Bladder

51715 Endoscopic injection/implant
52000 Cystoscopy
52001 Cystoscopy, removal of clots
52005 Cystoscopy & ureter catheter
52007 Cystoscopy and biopsy
52010 Cystoscopy & duct catheter
52204 Cystoscopy
52282 Cystoscopy, implant stent
52327 Cystoscopy, inject material
52330 Cystoscopy and treatment
52351 Cystouretro & or pyeloscope
52352 Cystouretro w/stone remove
52353 Cystouretero w/lithotripsy
52354 Cystouretero w/biopsy
52355 Cystouretero w/excise tumor
52402 Cystourethro cut ejacul duct
Surgery/Female Genital System: Cervix Uteri
57452 Examination of vagina
57454 Vagina examination & biopsy
57455 Biopsy of cervix w/scope
57456 Endocerv curettage w/scope
57460 Cervix excision
57461 Conz of cervix w/scope, leep
Surgery/Female Genital System: Corpus Uteri
58555 Hysteroscopy, dx, sep proc
58558 Hysteroscopy, biopsy
58559 Hysteroscopy, lysis
58560 Hysteroscopy, resect septum
58562 Hysteroscopy, remove fb
58565 Hysteroscopy, sterilization
Surgery/Female Genital System: Oviduct/Ovary
58670 Laparoscopy, tubal cautery
58671 Laparoscopy, tubal block
Surgery/Eye and Ocular Adnexa: Anterior Segment
66820 Incision, secondary cataract
66821 After cataract laser surgery
66830 Removal of lens lesion
66982 Cataract surgery, complex
66983 Remove cataract, insert lens

Other Surgeries:

14301 Skin Tissue Rearrangement

21011 Exc Face Les Sc< 2 cm

21012 Exc Face Les Sc=2 cm

21013 Exc Face Tum Deep < 2 cm

21014 Exc Face Tum Deep = 2 cm

21552 Exc Neck Les Sc = 3 cm

21554 Exc Neck Tum Deep = 5 cm

21558 Resect Neck Tum = 5 cm

21931 Exc Back Les Sc = 3 cm

21932 Exc Back Tum Deep < 5 cm

21933 Exc Back Tum Deep = 5 cm

22901 Exc Back Tum Deep = 5 cm

22902 Exc Abdomen Les Sc < 3 cm

22903 Exc Abdomen Les Sc > 3 cm

23071 Exc Shoulder Les Sc > 3 cm

23073 Exc Shoulder Tum Deep > 5 cm

24071 Exc Arm/Elbow Les Sc = 3 cm

24073 Exc Arm/Elbow Tum Deep > 5 cm

25071 Exc Forearm Les Sc > 3 cm

25073 Exc Forearm Tum Deep = 3 cm

26111 Exc Hand Les Sc > 1.5 cm

26113 Exc Hand Tum Deep > 1.5 cm

27043 Exc Hip Pelvis Les Sc > 3 CM

27045 Exc Hip/Pelvis Tum Deep > 5 CM

27337 Exc Thigh/Knee Les Sc > 3 CM

27339 Exc Thigh/Knee Tum Deep >5CM

27632 Exc Leg/Ankle Les Sc > 3cm

27634 Exc Leg/Ankle Tum Deep >5 cm

28039 Exc Foot/Toe Tum Sc > 1.5 cm

28041 Exc Foot/Toe Tum Deep >1.5cm

29581 Apply Multilay Comprs Lower Leg

31626 Bronchoscopy w/ Markers

32552 Remove Lung Catheter

36147 Access AV Dial Grft for Eval

36148 Access AV Dial Grft for Proc

37761 Ligate Leg Veins Open

51727 Cystometrogram w/UP

51728 Cystometrogram w/VP

51729 Cystometrogram w/VP&UP

53855 Insert Prost Uretheral Stent

63661 Remove Spine El Trd Perq Aray

63662 Remove Spine El Trd Plate

63663 Revise Spine El Trd Perq Aray

63664 Revise Spine El Trd Plate Revised

64490 Inj Paravert F Jnt C/T 1 LEV

64493 INJ Paravert F JNT L/S 1 LEV

0213T US Facet JT INJ CERV/T 1 LEV

0216T US Facet JT INJ LS 1 LEVEL

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

Number of stress SPECT MPI, stress echo, CCTA, and CMR performed

**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).*

All consecutive stress SPECT MPI, stress echocardiography, CCTA, and CMR orders

Measurement Entity: Imaging laboratory prospectively measured on test requisition forms and/or patient charts

Level of Measurement/Analysis: Imaging laboratory*

*Attribution for not appropriate use is shared between the ordering physician and imaging laboratory. In an ideal world, attribution to the ordering physician or institution, as well as the imaging laboratory, would be reflected in the reporting of these measures. However, there are numerous complexities that prevent assignment of these measures to individual ordering physicians. For example, ordering volumes from individual physicians and institutions are insufficient to make meaningful comparisons to allow such attribution. Thus, these measures will be reported at the level of the imaging laboratory. However, the extent to which the institution housing the imaging laboratory can impact these measures will be dependent upon cooperation of ordering physicians with the imaging laboratory.

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

None.

**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.)*

None.

**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.)*

None

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

No risk adjustment or risk stratification

If other:

**S.12. Type of score:**

Rate/proportion

If other:

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

Better quality = 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.)*

Locate all stress SPECT MPI, stress echocardiography, CCTA, and CMR orders performed during the sampling period.

Record the total number of tests during the sampling period as the denominator.

From this sets of test orders, identify orders containing the criteria listed in the numerator

**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.

Measures are to be developed based on a sample of a full calendar year based on the following sampling methodology:

Select a starting month:

- o January
- o March
- o May
- o July
- o September
- o November

Begin 60 day data collection period on the 1st on the month for the selected starting month

Determine whether at least 30 stress SPECT and stress echo orders have been placed during the selected time period. If not, select another time period with a minimum number of 30 cases. If no time period includes the minimum number of cases, then the imaging laboratory does not have sufficient volume to report this measure.

Sampling is required for this measure as full year data collection does not alter performance rates for this measure and would place an additional data collection burden on laboratories. It also allows laboratories to share performance with ordering physicians more quickly than would be possible under full year calendar reporting.

**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.

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

*If other, please describe in S.18.*

Other, Registry Data

**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.

Optimization of Patient Selection for Cardiac Imaging

**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)*

Available in attached appendix at A.1

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

Clinician : Group/Practice, Facility

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

Outpatient Services

If other:

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

**2. Validity – See attached Measure Testing Submission Form**

nqf_testing_attachment_7.1_670_July_2018_-_updated.docx

**2.1 For maintenance of endorsement**

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

No

**2.2 For maintenance of endorsement**

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

No

**2.3 For maintenance of endorsement**

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

No - This measure is not risk-adjusted.

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

**Measure Number** (*if previously endorsed*)**:** 670
**Measure Title**: Cardiac stress imaging not meeting appropriate use criteria: Preoperative evaluation in low risk surgery patients
**Date of Submission**: 11/1/2018

**Type of Measure:**

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

**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: Click here to describe | ☐ other: Click here to describe |

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

N/A

**1.3. What are the dates of the data used in testing**? August 15, 2007 and May 15, 2010

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

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

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

11 practices encompassing 12 ZIP codes within the Chicago metropolitan area; 20 primary care physicians and 2 cardiologists

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.
Circulation. 2013 Oct 8;128(15):1634-43. doi: 10.1161/CIRCULATIONAHA.113.002744. Epub 2013 Sep 10.

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

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Circulation. 2013 Oct 8;128(15):1634-43. doi: 10.1161/CIRCULATIONAHA.113.002744. Epub 2013 Sep 10

Baseline Clinical and Imaging Characteristics

| | Overall Cohort (n=1511) |
|---|---|
| Age, y | 59±13 |
| Women, n (%) | 657 (43.5) |
| Primary indication for MPI, n (%) | |
| Chest pain | 688 (45.5) |
| Dyspnea | 158 (10.5) |
| Abnormal ECG | 136 (9.0) |

| | Overall Cohort (n=1511) |
|---|---|
| Evaluation of known CAD | 159 (10.5) |
| Preoperative assessment | 37 (2.4) |
| Syncope | 21 (1.4) |
| Asymptomatic | 262 (17.3) |
| Hypertension, n (%) | 841 (55.6) |
| Diabetes mellitus, n (%) | 333 (22.0) |
| Dyslipidemia, n (%) | 695 (46.0) |
| Tobacco use, n (%) | 181 (12.0) |
| Family history of CAD, n (%) | 544 (36.0) |
| Framingham 10-y CHD risk, % | 13±10 |
| Likelihood of obstructive CAD, %* | 18±13 |
| Exercise stress (Bruce) protocol, n (%) | 1164 (77.0) |
| BMI, kg/m$^2$ | 30±5.7 |
| Known CAD, n (%) | 271 (17.9) |
| Previous CABG, n (%) | 76 (5.0) |
| Previous PCI, n (%) | 87 (5.8) |
| Previous MI, n (%) | 37 (2.4) |
| Statin, n (%) | 580 (38.4) |
| Antiplatelet, n (%) | 370 (24.5) |
| β-Blocker, n (%) | 307 (20.3) |
| ACE-I or ARB, n (%) | 567 (37.5) |
| Myocardial perfusion, n (%) | |
| Normal (SSS=0–3) | 1344 (88.9) |
| Mildly abnormal (SSS=4–8) | 79 (5.2) |
| Moderately abnormal (SSS=9–13) | 47 (3.1) |
| Severely abnormal (SSS >13) | 41 (2.7) |

| | Overall Cohort (n=1511) |
|---|---|
| Myocardial ischemia, n (%) | |
| None (SDS ≤1) | 1399 (92.6) |
| Mild (SDS=2–4) | 38 (2.5) |
| Moderate (SDS=5–7) | 40 (2.6) |
| Severe (SDS >7) | 43 (2.8) |
| Type of perfusion abnormality, n (%) | |
| Reversible | 87 (5.8) |
| Fixed | 61(4.0) |
| Reversible and fixed | 19 (1.3) |
| Poststress LVEF <50%, n (%) | 78 (5.2) |

**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 cohort used for the validity testing is described above. A smaller single center study was used for reliability testing and is cited below.

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging.

The demographics of the single center study are as follows: May 1, 2005, and May 15, 2005. Mayo Clinic (Rochester, Minn). The mean±SD age of the 298 study patients was 66±13 years; 52% were men, 20% had diabetes mellitus, 60% had hypertension, 66% had hyperlipidemia, 54% had a history of smoking, 11% had a prior myocardial infarction, 20% had prior coronary revascularization, 36% had chest pain, 38% had dyspnea, and 41% had a normal resting ECG.

**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.

N/A

---

**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*)

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging.

Using the appropriateness criteria document, 2 experienced cardiac registered nurse abstractors reviewed patient demographics and other relevant information and classified each patient as appropriate, inappropriate, or uncertain. Patients who did not fit any of the clinical situations in the appropriateness criteria were judged to be not classifiable. The level of agreement between the 2 raters was analyzed. Patients who did not fit the measure were deemed unclassified as they did not conform to the available scenarios. It does not imply that data was unavailable to determine the appropriateness of scenarios that had been published, including the focus of this measure.

Also, see section 2b1 for validity testing of data elements.

**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*)

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging. 2009 May;2(3):213-8. Nurse abstracter agreement kappa=0.72 for stress echocardiography

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

The data elements required for calculation of the appropriate use metrics can be obtained reliably by clinical staff from data residing in patient records with a high degree of agreement between nurses who would enter the data into the registry/clinical database.

---

**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)*

Relationship between appropriate use score and predictive value of SPECT MPI

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

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Patients in the appropriate/uncertain group experienced significantly higher overall rates of death (HR, 2.9; 95% CI, 1.05–8.0; P=0.04), the composite of death or MI (HR, 1.04; 95% CI, 1.01–1.07; P=0.03), and the composite of cardiac death or MI (HR=5.7; 95% CI, 1.3–25.6; P=0.02) after adjustment for clinical covariates. Among patients in the appropriate/uncertain group, abnormal MPI continued to predict a multifold increase in the risk of death, cardiac death, composite of death or MI, and composite of cardiac death or MI (Figure 4). However, in the inappropriate group, there were no statistically significant differences in MACE rates between subjects with abnormal versus normal MPI (Figure 4). Furthermore, using Cox regression models, no interaction was identified between the study group and MPI finding in predicting death, the composite of death or MI, or the composite of cardiac death or MI (P=0.91, 0.70, and 0.43, respectively).

A Cox regression model demonstrated that inappropriate MPI use was a negative predictor of all-cause mortality (HR, 0.26; 95% CI, 0.10–0.67; P=0.005) after adjustment for myocardial perfusion finding (normal versus abnormal; HR, 2.5; 95% CI, 1.1–5.9; P=0.04) and depressed LVEF (<50%; HR, 3.7; 95% CI, 1.5–9.3; P=0.006); undergoing early coronary revascularization was not predictive of mortality (P=0.98). Similarly, in separate models, we demonstrated that inappropriate use was an independent negative predictor of the secondary end points of death or MI (HR, 0.31; 95% CI, 0.14–0.70; P=0.005) and cardiac death or MI (HR, 0.16; 95% CI, 0.04–0.71; P=0.02) after adjustment for depressed LVEF, myocardial perfusion findings, and early revascularization. In these models, MPI and depressed LVEF independently predicted the composite end points of death or MI and cardiac death or MI, whereas undergoing early coronary revascularization after MPI was not predictive of these end points (P≥0.97). Finally, in forward stepwise Cox regression models, appropriate use was shown to have incremental prognostic value to perfusion imaging and depressed LVEF in predicting MACE; undergoing early revascularization (<60 days) did not provide significant additional predictive value

**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?*)

Appropriateness of imaging as measured by these metrics is correlated with the downstream value of the test in contributing to clinical decision making. As such, the metrics contribute to ensuring the prognostic value of the imaging procedures measured.

---

**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*)

**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*)

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

---

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

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

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **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.**

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

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

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

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

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

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

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

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

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

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

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

**2b3.9. Results of Risk Stratification Analysis**:

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

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

---

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

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

To date, there has been consistency in the studies showing the gap in performance on these metrics across sites. While individual practitioner level measurement and rates based on type of physician have shown variability, practice/hospital performance has been similar at baseline and after intervention to improve.

**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*)

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

No statistical tests have been applied to demonstrate differences among the measured entities at the practice/hospital level. Inappropriate use rates for individual practitioners ranged from 10% to 77% (P<0.001) and were higher among primary care physicians than cardiologists (47% versus 28%; P<0.001)

Fonseca R, Negishi K, Otahal P, et al. Temporal Changes in Appropriateness of Cardiac Imaging. J Am Coll Cardiol. 2015 Mar 3;65(8):763-73.

A separate meta-analysis demonstrated wide variation of appropriate use rates as described in the performance scores over time in section 2.b.1.

**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?*)

Inappropriate use is common among a wide range of practices and hospitals. Variability exists within practices and provides opportunities for peer to peer learning and improvement on these measures, especially within a practice or between primary care and specialists.

---

**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*)

**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*)

**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*)

**_____**

**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*)

All subjects were classifiable according to the 2009 AUC and therefore no analysis for missing data was required. For validity testing, some subjects were lost to follow-up. Their demographic and AUC patterns were analyzed for similarity to the included patient cohort.

**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*)

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Compared with subjects with complete follow-up, the patients excluded (n=182) or lost to follow-up (n=14) were younger (mean age, 55±15 versus 59±13 years; P=0.001) and had lower likelihood of obstructive CAD (15±13% versus 18±13%; P=0.007) but similar mean 10-year Framingham coronary heart disease risk (12.7±10.8% versus 12.8±10%; P=0.88) and CAD prevalence (19% versus 18%; P=0.62). The prevalence of depressed LVEF and abnormal perfusion was nearly identical (P=0.97 and 0.89, respectively), with a similar breakdown of reversible, fixed, and mixed defects (P=0.64). The excluded patients had a similar distribution of AUC classifications: 104 (53.1%) appropriate, 89 (45.4%) inappropriate, and 3 (1.5%) uncertain (P=0.53).

**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*)

Performance results were not biased as no missing data was recorded for the metrics themselves. Validity testing showed similar distribution of AUC, perfusion defects, and CAD prevalence and thus unlikely to have impacted results of this testing.

## 3. Feasibility

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

**3a. Byproduct of Care Processes**

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

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

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Other

If other: Decision Support Tool feeding registry

**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**.

Some data elements are in defined fields in electronic sources

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

Some data elements should already be a part of the electronic record (PCI history, scheduled surgery). In addition, e-ordering for diagnostic testing has been proposed for meaningful use, encouraging integration of these types of data elements. In addition, ACC is developing clinical decision support tools that can be embedded in electronic health records to capture the necessary information.

**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.**

Hendel, RC; Cerqueira, M; Douglas, PS et al. "A Multicenter Assessment of the Use of Single-Photon Emission Computed Tomography Myocardial Perfusion Imaging With Appropriateness Criteria". J Am Coll Cardiol. Published online December 10, 2009.

This study demonstrated the feasibility of data collection as well as the most frequent inappropriate indications. This allowed ACC to narrow the number of indications measured for this measure set along with the associated data elements.

**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)*.

None required. Decision support tools are available to aid in data collection and are available on a per test basis.

## 4. Usability and Use

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

**4a. Accountability and Transparency**

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

**4.1. Current and Planned Use**

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

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|

| | |
|---|---|
| | Public Reporting<br>PQRS<br>http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html<br>Payment Program<br>Quality Payment Program<br>https://qpp.cms.gov/mips<br>Quality Payment Program<br>https://qpp.cms.gov/mips<br>Regulatory and Accreditation Programs<br>IAC<br>http://www.intersocietal.org/intersocietal.htm<br>Professional Certification or Recognition Program<br>FOCUS<br>www.cardiosource.org/focus<br>Quality Improvement (Internal to the specific organization)<br>FOCUS<br>https://www.acc.org/focus |

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

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

MIPS - CMS/pay for performance/national; The data collected at the lab level for this measure can be further segmented by physician to help them understand their appropriate use patterns; although small sample sizes can limit comparability for some providers
FOCUS - ACC/lab accreditation, quality improvement and utilization management/national - 25,000 cases with concentrations in DE (100% for SPECT MPI) and Western PA (10% for SPECT MPI and stress echo for cardiologists) - addtional 6,000 cases
IAC - lab accreditation/national - 100% - 5% of lab tests performed on an annual basis

**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.**

through CMS

**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.**

through CMS

**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.**

through CMS

**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.**

In a cohort study of a 5% national sample of Medicare beneficiaries, annual rates of overall testing appeared to increase from 2000 to 2008 and then declined until 2016. Rates of low-value tests (preoperative stress testing and routine stress testing after coronary revascularization) appeared to have increased and then decreased.

**4b2. Unintended Consequences**

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

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

There is not an comprehensive list of surgeries that are low risk. Both a definition and example list of surgeries is provided, but clinical judgment also is needed.

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

## 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)**

0669 : Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac, Low Risk Surgery

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

**5a. Harmonization of Related Measures**

The measure specifications are harmonized with related measures;

**OR**

The differences in specifications are justified

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

**Are the measure specifications harmonized to the extent possible?**

No

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

Different populations and data sources used

**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.)**

This measure provides an additional level of analysis that applies not only to hospitals but also outpatient physician clinics. The data source also provides a richer source of clinical information to distinguish between testing ordered for preoperative assessment and other cardiovascular causes co-existing at the same time.

## Appendix

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

Attachment **Attachment:** FOCUS_Data_Collection_Sheet-635249619633321013.docx

## Contact Information

**Co.1 Measure Steward (Intellectual Property Owner):** American College of Cardiology

**Co.2 Point of Contact:** Amy, Dearborn, adearborn@acc.org, 202-375-6257-

**Co.3 Measure Developer if different from Measure Steward:** American College of Cardiology Foundation

**Co.4 Point of Contact:** Joseph, Allen, jallen@acc.org, 202-375-6463-

## 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.**

All individuals are volunteer members representing American College of Cardiology Foundation:

Pamela Douglas, MD, MACC

Joseph Allen, MA

Robert Hendel, MD, FACC

Joseph Cacchione, MD, FACC

Manuel Cerqueira, MD, FACC

Joseph Drozda, MD, FACC

Michael Picard, MD, FACC

Martha Radford, MD, FACC

Leslee Shaw, PhD, FACC

Allen Taylor, MD, FACC

Group developed list of proposed measures, specifications, definitions, justification, etc.

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 2009

**Ad.3 Month and Year of most recent revision:** 11, 2019

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

**Ad.5 When is the next scheduled review/update for this measure?** 11, 2019

**Ad.6 Copyright statement:** Copyright 2013. American College of Cardiology Foundation

**Ad.7 Disclaimers:** date above refers to maintenance of endorsement

**Ad.8 Additional Information/Comments:** date above refers to maintenance of endorsement

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 #:** 0671

**Corresponding Measures:**

**De.2. Measure Title:** Cardiac stress imaging not meeting appropriate use criteria: Routine testing after percutaneous coronary intervention (PCI)

**Co.1.1. Measure Steward:** American College of Cardiology

**De.3. Brief Description of Measure:** Percentage of all stress SPECT MPI, stress echo, CCTA and CMR performed routinely after PCI, with reference to timing of test after PCI and symptom status.

**1b.1. Developer Rationale:** Appropriate use criteria define "when to do" and "how often to do" a given procedure in the context of scientific evidence, the health care environment, the patient's profile and a physician's judgment. While practice guidelines provide a foundation for summarizing evidence-based cardiovascular care or for providing expert consensus opinions, in many areas, marked variability remains in the use of cardiovascular procedures, raising questions about over-use and under-use. Appropriate use criteria provide practical tools to measure this variability and to look at utilization patterns. The criteria are designed to examine the use of diagnostic and therapeutic procedures to support efficient use of medical resources, while also providing patients with quality, appropriate care.

A measure that reports rates of inappropriate imaging within practices would contain information regarding both cost and quality, because an inappropriate test results in both higher costs and poorer-quality care. Conversely, a reduction in this rate would simultaneously improve quality and decrease cost. Improvements in this metric should lead to consistent application of AUC and improve the efficiency of the system.

**S.4. Numerator Statement:** Number of stress SPECT MPI, stress echo, CCTA and CMR performed in asymptomatic patients within 2 years of the most recent PCI

**S.6. Denominator Statement:** Number of stress SPECT MPI, stress echo, CCTA and CMR performed

**S.8. Denominator Exclusions:** None

**De.1. Measure Type:** Efficiency

**S.17. Data Source:** Other, Registry Data

**S.20. Level of Analysis:** Clinician : Group/Practice, Facility

**IF Endorsement Maintenance – Original Endorsement Date:** Apr 26, 2011 **Most Recent Endorsement Date:** Jun 29, 2015

## Preliminary Analysis: Maintenance of Endorsement

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

### Criteria 1: Importance to Measure and Report

#### 1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

**Evidence Summary**

- This measure is an "appropriate use" measure – a type of process measure – rather than a true efficiency measure that would include the quality and cost components. In the measure rationale, the developer states this measure concerns both cost and quality, as inappropriate tests result in high costs and poor care quality.
- The developer provided evidence from the *2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the Diagnosis and Management of Patients With Stable Ischemic Heart Disease*. The guidelines note that "Nuclear MPI, echocardiography, or CMR, with either exercise or pharmacological stress or CCTA, is not recommended for follow-up assessment in patients with SIHD, if performed more frequently than … 2-year intervals after PCI."
    - The evidence was assigned "C" grade indicating "Very limited patient populations evaluated. Only consensus opinion of experts, case studies, or standard of care" The recommendation was assigned "Class III: No Benefit" grading, which corresponds to "conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective, and in some cases may be harmful."
    - The QCC included with the guideline indicates no studies existed that examined outcomes for patients undergoing imaging post-PCI (at the time of guideline publishing).
- The developer identified six studies examining the outcomes of patient post-PCI who underwent imaging. These studies were published after the guideline. No QCC is included for the studies. It is difficult to evaluate the quality of the studies based on what is presented. They appear to be observational studies, and most were examining the association between imaging and repeat revascularization.

**Changes to evidence from last review**

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

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

**Updates:** N/A

**Exception to evidence**

N/A

***Questions for the Committee:***

- How strong is the evidence for this relationship?
- Is the evidence directly applicable to the process of care being measured?

For possible exception to the evidence criterion:

- Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?
- Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?
- Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?

**Guidance from the Evidence Algorithm**

Process measure based on systematic review (Box 3) → Guidelines based on expert opinion (Box 7) → No empirical evidence (Box 10) → INSUFFICIENT

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

**RATIONALE:** The guidelines presented represent expert opinions in an area where few to no studies exist. The additional studies do not provide enough information to assess study quality and are weakly linked to the measure focus.

## 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.

- For maintenance measures, performance scores on the measure as specified at the specified level of analysis are required for maintenance of endorsement. The results provided do not appear to be calculated using the measure as specified. They do not include the same tests as the measure, and it is difficult to tell if the calculations were performed in alignment with the measure specifications.
- The developer presented site-specific performance score, which were obtained from a sub-analysis of the data collected for one study. The study is from 2010.
    - Six sites participated in the pilot study including 3 urban, 2 suburban, and 1 rural location in Florida, Wisconsin, Oregon, and Arizona. The number of cardiologists at each site ranged from 7 to 20 physicians. The number of SPECT MPI patients submitted from each site varied from 328 to 1,597 patients, but a total of 6,351 subjects with complete data were entered into the pilot database.
    - The developer provided results for four sites with results ranging from 0.9% to 4.8%. No specific information is provided about each of the site, i.e., size, number of studies, location, ownership, or the timeframe when the data were obtained.

    - There is not enough information to determine if the results provided correspond to the levels of analysis for which this measure is specified. The study only includes one of the four types of tests included in the measure.
- The developer summarizes three appropriateness studies as evidence for improvement and cites three additional articles.

**Disparities**

- Disparities data from the measure as specified is required for maintenance of endorsement. No disparities data is provided.
- The developer does provide information from literature identifying older patients and patients with some clinical risk factors as being less likely to receive downstream testing. Patients' socioeconomic status and hospitals' teaching status were associated with higher post-PCI testing. It is unclear if these associations are with any downstream testing or with testing identified as inappropriate in accordance with this measure.

***Questions for the Committee:***

- Does the developer provide enough data to show a gap in care that warrants a national performance measure?
- Does the data provided demonstrate a need for this measure?
- Since the developer did not provide any information on disparities, are you aware of evidence that disparities exist in this area of healthcare?

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

**RATIONALE:** The data provided for performance gap and disparities is minimal or insufficient. The data provided are from 2010, providing no information on current performance gaps. Performance scores on the measure as specified are required for maintenance of endorsement. Those scores are not provided.

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

1a.

- (I had to start over because I ran out of time). 0670, 0671, and 0672 address the same problem--overuse. This is a different task than that of the vast majority of NQF-endorsed measures--assessing underuse. So, I'm not convinced that the template to evaluate underuse measures is useful for our task. I think the evidence is insufficient with exception
- Interesting conundrum. We approve positive quality measures when there is high quality evidence they improve outcomes. This is an example of an expensive test with some risk that has been/is being commonly done with no evidence for benefit. Lack of evidence may be a result of positive publication bias.
- Old guideline based on level of evidence C
- There is a 2014 focused update on the guidelines. There is insufficient evidence. Agree with staff notes.
- low to insufficient
- No new evidencd

1b.

- The performance gap published in a 2010 paper ranged from 0.9% to 4.8%. No newer data were provided.
- no data to evaluate
- 2010 data from small sample
- Majority of data from a 2010 study. No gap analyzed from current data. Agree with NQF Staff
- low to insufficient

- no new data since 2010

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability: Specifications and Testing**

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

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

### Reliability

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

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

### Validity

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

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

**Composite measures only:**

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

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

**Evaluators:** NQF Staff

Scientific Acceptability: Preliminary Analysis Form

**Measure Number:** 0671

**Measure Title:** Cardiac stress imaging not meeting appropriate use criteria: Routine testing after percutaneous coronary intervention (PCI)

**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: Unclear**

**Measure is:**

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

**RELIABILITY: SPECIFICATIONS**

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

   **Submission document:** "MIF_0671" 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.**
   - Does the measure include all ages? No age range is included in the specifications.
   - The developer indicates Clinician: Group/Practice is a level of analysis for this measure. It is unclear which clinician would be held accountable. The denominator of number of tests performed doesn't correspond to an ordering physician. Is it the performing physician? The attribution should be clear.

**RELIABILITY: TESTING**

**Submission document:** "MIF_0671" 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**
   - Information supplied in testing attachment does not appear to correspond to data source or levels of analysis indicated.

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☒ **No**

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

   **Submission document:** Testing attachment, section 2a2.2

   - The developer states reliability was tested at the data element level.
   - The study included to demonstrate reliability testing is a single-center study including 298 patients. It includes stress echocardiogram and SPECT MPI, but not the other cardiac tests included in the measure specifications.
   - The study included appears to focus on using appropriate use criteria to evaluate the appropriateness of a test whereas this measure attempts to identify tests used for monitoring after PCI. It's unclear how precisely the appropriate use criteria in the study correspond to the measure specifications.
   - The inter-rater reliability provided is for the level of agreement in two nurses' appropriateness ratings for the cardiac testing. Appropriateness ratings are not a data element of this measure.

The relationship between the appropriateness ratings and the measure specifications is unclear.

7. **Assess the results of reliability testing**
   - There is not enough information provided to assess the reliability of this measure or its data elements.

   **Submission document:** Testing attachment, section 2a2.3

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

   **Submission document:** Testing attachment, section 2a2.2

   ☐ **Yes**

   ☒ **No**

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

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

   **Submission document:** Testing attachment, section 2a2.2

   ☐ **Yes**

   ☒ **No**

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

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

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

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

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

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

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - There is not enough information provided to assess the reliability of this measure or its data elements. The information provided in the reliability section is not clearly related to the measure score or to the data elements in the measure. Testing does not appear to correspond to the levels of analysis (clinician: group/practice and facility) indicated for the measure.
    - In addition to concerns with the testing, staff identified concerns with the clarity of the specifications, particularly clinician attribution.

### VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

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

    - The developer indicates that there are no exclusions for this measure.

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

    **Submission document:** Testing attachment, section 2b4.

- The developer's discussion of differences in performance focuses on inappropriate use and it is unclear if these results are in line with the focus of this measure or a more general application of AUC.
- The developer provides no details on statistical testing of measure results.
- While the developer notes that there is variation in inappropriate use rates at the individual-practitioner level and that these rates vary by physician specialty, no method is highlighted to identify meaningful differences in performances.
- Previously one of the committee members questioned whether the general data supplied proves reliability.

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.
    - Not applicable

15. **Please describe any concerns you have regarding missing data.**
    **Submission document:** Testing attachment, section 2b6.
    - The developer reports "All subjects were classifiable according to the 2009 AUC and therefore no analysis for missing data was required." It is unclear which patients are being referenced and the relationship between the data in the study and the data elements of this measure is unclear.

16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

**<u>For cost/resource use measures ONLY</u>:**

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

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

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

## VALIDITY: TESTING

19. **Validity testing level: ☐ Measure score ☐ Data element ☐ Both**
20. **Method of establishing validity of the measure score:**
    - ☐ **Face validity**
    - ☐ **Empirical validity testing of the measure score**
    - ☒ **N/A (score-level testing not conducted)**
21. **Assess the method(s) for establishing validity**

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

    - The developer states their method of validity testing is the "relationship between appropriate use score and predictive value of SPECT MPI." This does not appear to align with face or empirical validity testing for measure 0671.

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

    **Submission document: Testing attachment, section 2b2.3**

    - The results presented do not provide information that can be used to assess the validity of this measure. There is not enough relevant information provided.

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

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

    ☐ **Yes**

    ☒ **No**

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

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

    *NOTE that data element validation from the literature is acceptable.*

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

    ☐ **Yes**

    ☐ **No**

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

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

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

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

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

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

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

The information provided in the validity section is not directly related to the measure score or to the measure's data elements. There is not enough information provided to assess the validity of the measure score or the data elements.

### ADDITIONAL RECOMMENDATIONS

27. **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.**

***Questions for the Committee regarding reliability:***

- Is it clear from the provided specifications how this measure would be attributed to a clinician group or practice and which clinician group or practice would be held accountable?
- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The staff was not satisfied with the reliability testing for the measure. Does the Committee agree with the staff assessment?

***Questions for the Committee regarding validity:***

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The staff was not satisfied with the validity analyses for the measure. Does the Committee agree with the staff assessment?

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

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

2a1.

- PCI CPT codes are not provided
- audit for data may be present, but how they apply to measure is not clear
- No mention of age limits
- Agree with NQF assessment of the reliability
- the major concern is which physician is being held accountable (the ordering or performing or both)
- difficult to follow

2a2.

- The measure relies on publications that are 10 years old.
- agree with insufficient evidence
- Based on single center study of ~300 patients
- Agree with NQF assessment of the reliability
- yes, there could be difficulty ascertaining both the accurate of stress tests in the numerator and the denominator given that tests could be performed outside the group/single EMR
- lacking appropriate methods and results

2b1.

- In Doukky, patients in the uncertain and appropriate categories had higher event rates than patients in the inappropriate group.
- not enough information
- None provided
- Agree with NQF assessment of the validity
- no concerns
- not sure about the testing

2b4-7.

- Representativeness of the patient sample could be wuestioned.
- cannot judge
- No testing provided
- Agree with NQF assessment of the validity
- the definition of "all tests ordered" is difficult to implement
- unclear about meaningful differences

2b2-3.

- The analyses are risk adjusted and I accept that it is valid
- unclear how ASX is decided for numerator
- n/a
- Agree with NQF assessment of the validity
- there is inadequate data provided
- no risk adjustment

## Criterion 3. Feasibility

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

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

- The developer suggests that not all data elements are in defined fields in electronic sources, though some should already be part of the EHR (e.g., PCI history, scheduled surgery).
- The developer cites Hendel et al. as a source demonstrating the feasibility of data collection. This study also demonstrated the most frequent inappropriate indications, used by the developer to narrow the number of indications measured for this measure set, along with the associated data elements.
- The developer states no fees, licensing, or other requirements to use any aspect of this measure are required. Decision support tools are available to aid on a per test basis.

***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?
- What is the burden of data collection, i.e., chart abstraction and data entry to a registry?

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

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

- I agree that feasibility is moderate
- Unclear as to source. EHR, paper, registry, claims?
- no concerns
- The data seem vague - either with how it is captured or obtained digitally. Not specific.
- it is feasible if integrated within the EMR
- Not sure if completely specified by EHR

## Criterion 4: Usability and Use

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

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

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

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

**Current uses of the measure**

**Publicly reported?** ☒ **Yes** ☐ **No**

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

**OR**

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

**Accountability program details**

- QPP/MIPS
- FOCUS
- IAC

**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:** Developer states, "through CMS."

**Additional Feedback:** N/A

***Questions for the Committee:***

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

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

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

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

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

**Improvement results**

- Developer sites Luca SR, et al. and states that the authors observed a decrease in the use of stress testing after PCI procedures over time.
- Trends of results for this specific measure were not provided.

**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 reports there have been no unexpected findings.

**Potential harms**

- According to the developers, no unintended consequences have been identified for this measure.

**Additional Feedback:**

- None reported by developer.

***Questions for the Committee:***

- Are you aware of any unintended consequences for this measure?
- Do the benefits of the measure outweigh any potential unintended consequences?

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

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

4a1.

- The measure is publicly reported and currently used.
- Being used, but no transparency
- Yes in MIPS
- Good use in publicly reported programs
- feedback is mainly financial/CMS
- used in a publicaly reported program

4b1.

- In Ontario, The 2-year rate of stress testing declined significantly, from 68.1% among patients who underwent PCI in 2004 to 60.4% in 2012 ($p < 0.001$). (Luca, CMAJ Open) This is some progress in a Canadian province. US data are not presented.
- Unlikely to be harms
- No concerns
- Cite a study but does not give a summary

- none known
- useable

## Criterion 5: Related and Competing Measures

**Related or competing measures**

- Developer states there are no related or competing measures; however, NQF has identified the following related measures:
  - 0669 Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac Low-Risk Surgery (CMS)
  - 0670 Cardiac stress imaging not meeting appropriate use criteria: Preoperative evaluation in low risk surgery patients (ACCF)
  - 0672 Cardiac stress imaging not meeting appropriate use criteria: Testing in asymptomatic, low risk patients (ACCF)

**Harmonization**

- None provided by the developer.

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

- 0670, 0671 and 0672 are all appropriate use measures
- Related to other overuse measures
- 0672: no harmonization information provided
- None that are mentioned.
- None
- Respondent skipped this question

## Public and Member Comments

**Comments and Member Support/Non-Support Submitted as of: January 21, 2020**

- **No NQF members have submitted a support/non-support choice as of this date.**

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

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

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

NQF_evidence_attachment_Sep2017_-_671.docx

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

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

No

### 1a. Evidence

**Measure Number** (*if previously endorsed*)**:** 671

**Measure Title**: Cardiac stress imaging not meeting appropriate use criteria: Routine testing after percutaneous coronary intervention (PCI)

**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**: 11/7/2019

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

Outcome

☐ Outcome: Resource Use and Avoidance of Negative Clinical Benefit Risk Ratio for Patient Population

☐ 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: Resource Use and Avoidance of Negative Clinical Benefit Risk Ratio for Patient Population

☒ Appropriate use measure: Cardiac stress imaging not meeting appropriate use criteria: Routine testing after percutaneous coronary intervention (PCI)

☐ 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.

Measurement of appropriate use summarizes the financial value/resources use and avoidance of a negative clinical benefit risk ratio across a patient population in which a procedure is used. Various factors influence the ability of a procedure to contribute to the diagnosis and treatment of a patient, including the clinical factors summarized by appropriate use measures. These clinical factors combined with physician and patient decision making determine the probability that a procedure will have the intended impact on health outcomes of the patient

**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.**

**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

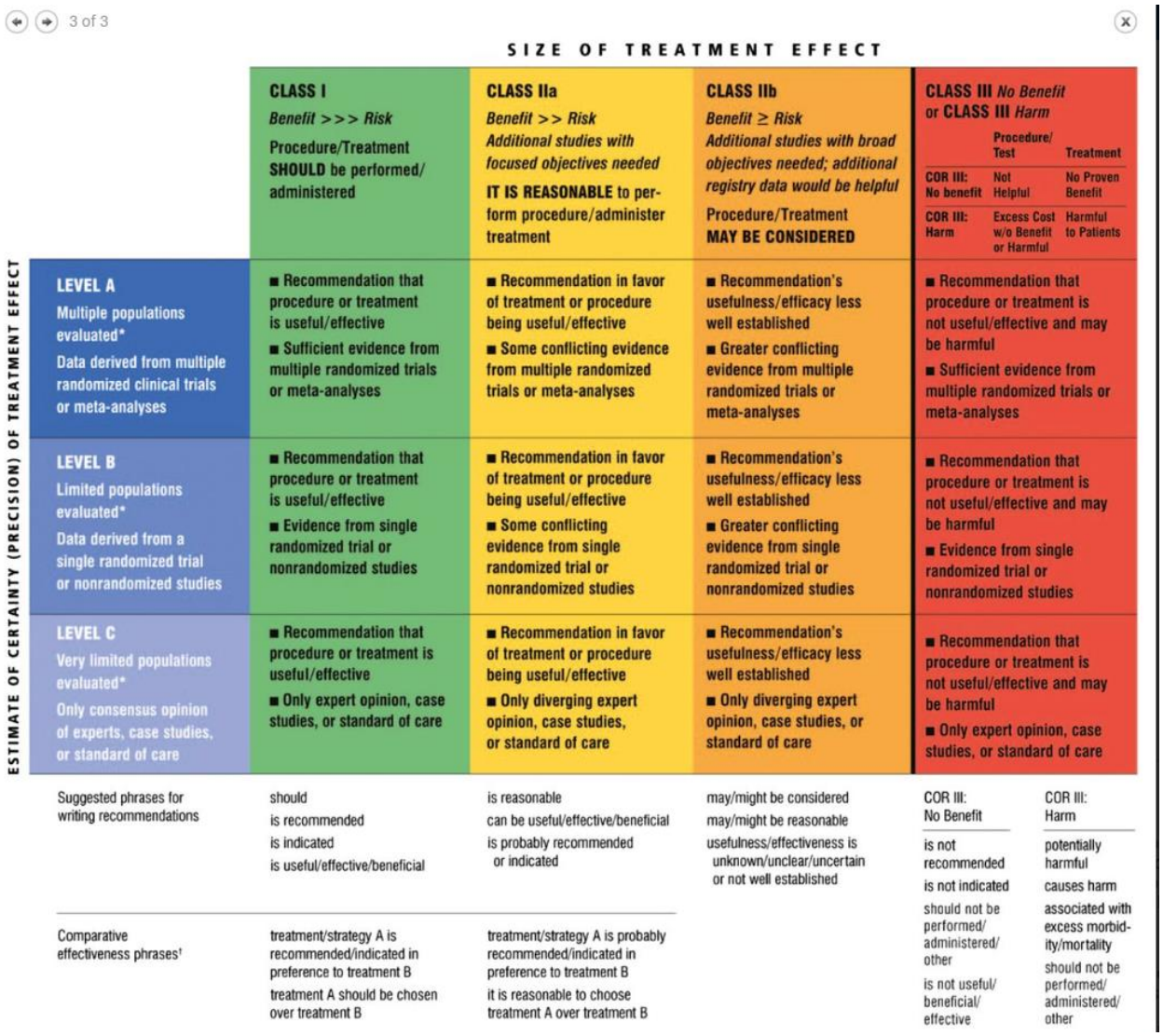
| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | **2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the Diagnosis and Management of Patients With Stable Ischemic Heart Disease: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons**<br><br>**Stephan D. Fihn, MD, MPH; Julius M. Gardin, MD; Jonathan Abrams, MD; Kathleen Berra, MSN, ANP; James C. Blankenship, MD; Apostolos P. Dallas, MD; Pamela S. Douglas, MD; Joanne M. Foody, MD; Thomas C. Gerber, MD, PhD; Alan L. Hinderliter, MD; Spencer B. King III, MD; Paul D. Kligfield, MD; Harlan M. Krumholz, MD; Raymond Y.K. Kwong, MD; Michael J. Lim, MD; Jane A. Linderbaum, MS, CNP-BC; Michael J. Mack, MD; Mark A. Munger, PharmD; Richard L. Prager, MD; Joseph F. Sabik, MD; Leslee J. Shaw, PhD; Joanna D. Sikkema, MSN, ANP-BC; Craig R. Smith, Jr, MD; Sidney C. Smith, Jr, MD; John A. Spertus, MD, MPH; Sankey V. Williams, MD**<br><br>**December 2012**<br><br>**J Am Coll Cardiol 2012;60:e44 –164** |

| | |
|---|---|
| | https://www.sciencedirect.com/science/article/pii/S0735109712027027?via%3Dihub |

| | |
|---|---|
| **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.** | **Page e124**<br>**CLASS III: No Benefit 1. Nuclear MPI, echocardiography, or CMR, with either exercise or pharmacological stress or CCTA, is not recommended for follow-up assessment in patients with SIHD, if performed more frequently than at a) 5-year intervals after CABG or b) 2-year intervals after PCI (10,12,15). (Level of Evidence: C)** |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | Level of Evidence C: Very limited patient populations evaluated. Only consensus opinion of experts, case studies, or standard of care. |
| Provide all other grades and definitions from the evidence grading system | See below* |
| Grade assigned to the **recommendation** with definition of the grade | CLASS III: NO BENEFIT<br>Conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective, and in some cases may be harmful. |
| Provide all other grades and definitions from the recommendation grading system | See below* |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | Quantity: No studies addressed outcomes of patient post PCI who underwent imaging among the guideline literature search. However, a number of studies have examined this question since the literature search was completed as detailed below.<br><br>Quality: n/a |

| **Estimates of benefit and consistency across studies** | **This measure looks at the absence of potential benefit in a specific population which is derivative of the studies examined but not a direct end point of the studies reviewed.** |
|---|---|
| What harms were identified? | The studies did not examine harm. |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | Shah BR, Cowper PA, O'Brien SM, et al. Patterns of cardiac stress testing after revascularization in community practice. J Am Coll Cardiol. 2010 Oct 12;56(16):1328-34<br>Although there is limited consensus as to the appropriate role of elective stress testing after coronary revascularization, more than one half of all patients in community practice had at least 1 stress test within 24 months of revascularization. Yield on such testing was low: only 5% of patients tested ultimately required repeat revascularization. These findings support the need to define better the role of stress testing after recent revascularization.<br><br>Shah BR, McCoy LA, Federspiel JJ, Mudrick D, Cowper PA, Masoudi FA, Lytle BL, Green CL, Douglas PS. Use of stress testing and diagnostic catheterization after coronary stenting: association of site-level patterns with patient characteristics and outcomes in 247,052 Medicare beneficiaries. J Am Coll Cardiol. 2013 Jul 30;62(5):439-46.<br>Although patient characteristics were largely independent of rates of post-PCI testing, higher testing rates were not associated with lower risk for myocardial infarction or death, but repeat revascularization was significantly higher at these sites. Additional studies should examine whether increased testing is a marker for improved quality of post-PCI care or simply increased health care utilization.<br><br>Mudrick DW, Shah BR, McCoy LA, at al. Patterns of stress testing and diagnostic catheterization after coronary stenting in 250 350 medicare beneficiaries. Circ Cardiovasc Imaging. 2013 Jan 1;6(1):11-9.<br>In this descriptive analysis, ST and invasive CA were common in older patients after PCI. Paradoxically, patients with higher risk features at baseline were |

| | |
|---|---|
| | less likely to undergo post-PCI testing. The revascularization yield was low on patients referred for ST after PCI, with only 7% [corrected] undergoing revascularization within 90 days.<br><br>Peterson T, Askew JW, Bell M, et al. Low yield of stress imaging in a population-based study of asymptomatic patients after percutaneous coronary intervention. Circ Cardiovasc Imaging. 2014 May;7(3):438-45.<br><br>In a population-based sample of patients undergoing PCI primarily for acute coronary syndromes, 1 in 8 had subsequent stress imaging when they were asymptomatic. These stress imaging tests resulted in further revascularization in <1% of patients. The low rate of downstream revascularization suggests that stress imaging in asymptomatic patients after PCI has low value.<br><br>Shah BR, Cowper PA, O'Brien SM, et al. Association between physician billing and cardiac stress testing patterns following coronary revascularization. JAMA. 2011 Nov 9;306(18):1993-2000.<br><br>Nuclear stress testing and stress echocardiography testing following revascularization were more frequent among patients treated by physicians who billed for technical fees, professional fees, or both compared with those treated by physicians who did not bill for these services.<br><br>Rossi JS, Federspiel JJ, Crespin DJ, et al. Stress imaging use and repeat revascularization among medicare patients with high-risk coronary artery disease. Am J Cardiol. 2012 Nov 1;110(9):1270-4.<br><br>Stress testing is commonly performed among Medicare patients after the initial revascularization, and most repeat procedures are performed for stable coronary artery disease. The variation in stress testing patterns only explained a modest fraction of the regional variation in the repeat revascularization rates. |



SIZE OF TREATMENT EFFECT

ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT

| | CLASS I<br>*Benefit >>> Risk*<br>Procedure/Treatment **SHOULD** be performed/ administered | CLASS IIa<br>*Benefit >> Risk*<br>*Additional studies with focused objectives needed*<br>**IT IS REASONABLE** to perform procedure/administer treatment | CLASS IIb<br>*Benefit ≥ Risk*<br>*Additional studies with broad objectives needed; additional registry data would be helpful*<br>Procedure/Treatment **MAY BE CONSIDERED** | CLASS III *No Benefit* or CLASS III *Harm*<br>Procedure/ Test / Treatment<br>COR III: No benefit — Not Helpful — No Proven Benefit<br>COR III: Harm — Excess Cost w/o Benefit or Harmful — Harmful to Patients |
|---|---|---|---|---|
| **LEVEL A**<br>Multiple populations evaluated*<br>Data derived from multiple randomized clinical trials or meta-analyses | ■ Recommendation that procedure or treatment is useful/effective<br>■ Sufficient evidence from multiple randomized trials or meta-analyses | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Some conflicting evidence from multiple randomized trials or meta-analyses | ■ Recommendation's usefulness/efficacy less well established<br>■ Greater conflicting evidence from multiple randomized trials or meta-analyses | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Sufficient evidence from multiple randomized trials or meta-analyses |
| **LEVEL B**<br>Limited populations evaluated*<br>Data derived from a single randomized trial or nonrandomized studies | ■ Recommendation that procedure or treatment is useful/effective<br>■ Evidence from single randomized trial or nonrandomized studies | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Some conflicting evidence from single randomized trial or nonrandomized studies | ■ Recommendation's usefulness/efficacy less well established<br>■ Greater conflicting evidence from single randomized trial or nonrandomized studies | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Evidence from single randomized trial or nonrandomized studies |
| **LEVEL C**<br>Very limited populations evaluated*<br>Only consensus opinion of experts, case studies, or standard of care | ■ Recommendation that procedure or treatment is useful/effective<br>■ Only expert opinion, case studies, or standard of care | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Only diverging expert opinion, case studies, or standard of care | ■ Recommendation's usefulness/efficacy less well established<br>■ Only diverging expert opinion, case studies, or standard of care | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Only expert opinion, case studies, or standard of care |
| Suggested phrases for writing recommendations | should<br>is recommended<br>is indicated<br>is useful/effective/beneficial | is reasonable<br>can be useful/effective/beneficial<br>is probably recommended or indicated | may/might be considered<br>may/might be reasonable<br>usefulness/effectiveness is unknown/unclear/uncertain or not well established | COR III: No Benefit: is not recommended; is not indicated; should not be performed/ administered/ other; is not useful/ beneficial/ effective<br>COR III: Harm: potentially harmful; causes harm; associated with excess morbidity/mortality; should not be performed/ administered/ other |
| Comparative effectiveness phrases† | treatment/strategy A is recommended/indicated in preference to treatment B<br>treatment A should be chosen over treatment B | treatment/strategy A is probably recommended/indicated in preference to treatment B<br>it is reasonable to choose treatment A over treatment B | | |

* grades and definitions from the evidence grading system

---

**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.*

Measurement of appropriate use summarizes the financial value/resources use and avoidance of a negative clinical benefit risk ratio across a patient population in which a procedure is used. Various factors influence the ability of a procedure to contribute to the diagnosis and treatment of a patient, including the clinical factors summarized by appropriate use measures. These clinical factors combined with physician and patient decision making determine the probability that a procedure will have the intended impact on health outcomes of the patient.

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Circulation. 2013 Oct 8;128(15):1634-43. doi: 10.1161/CIRCULATIONAHA.113.002744. Epub 2013 Sep 10.

Patients in the appropriate/uncertain group experienced significantly higher overall rates of death (HR, 2.9; 95% CI, 1.05–8.0; P=0.04), the composite of death or MI (HR, 1.04; 95% CI, 1.01–1.07; P=0.03), and the composite of cardiac death or MI (HR=5.7; 95% CI, 1.3–25.6; P=0.02) after adjustment for clinical covariates. Among patients in the appropriate/uncertain group, abnormal MPI continued to predict a multifold increase in the risk of death, cardiac death, composite of death or MI, and composite of cardiac death or MI (Figure 4). However, in the inappropriate group, there were no statistically significant differences in MACE rates between subjects with abnormal versus normal MPI (Figure 4). Furthermore, using Cox regression models, no interaction was identified between the study group and MPI finding in predicting death, the composite of death or MI, or the composite of cardiac death or MI (P=0.91, 0.70, and 0.43, respectively).

A Cox regression model demonstrated that inappropriate MPI use was a negative predictor of all-cause mortality (HR, 0.26; 95% CI, 0.10–0.67; P=0.005) after adjustment for myocardial perfusion finding (normal versus abnormal; HR, 2.5; 95% CI, 1.1–5.9; P=0.04) and depressed LVEF (<50%; HR, 3.7; 95% CI, 1.5–9.3; P=0.006); undergoing early coronary revascularization was not predictive of mortality (P=0.98). Similarly, in separate models, we demonstrated that inappropriate use was an independent negative predictor of the secondary end points of death or MI (HR, 0.31; 95% CI, 0.14–0.70; P=0.005) and cardiac death or MI (HR, 0.16; 95% CI, 0.04–0.71; P=0.02) after adjustment for depressed LVEF, myocardial perfusion findings, and early revascularization. In these models, MPI and depressed LVEF independently predicted the composite end points of death or MI and cardiac death or MI, whereas undergoing early coronary revascularization after MPI was not predictive of these end points (P≥0.97). Finally, in forward stepwise Cox regression models, appropriate use was shown to have incremental prognostic value to perfusion imaging and depressed LVEF in predicting MACE; undergoing early revascularization (<60 days) did not provide significant additional predictive value

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

Each of the documents below covers a clinical imaging procedure and was developed using the AUC methodology cited below.

The Appropriate Use Criteria have been published and updated on a regular basis by the American College of Cardiology in partnership with other societies and stakeholders. The evidence underlying the AUC appear in guidelines and systematic reviews contained in the appendix materials for these documents. The clinical indications and expert opinion used have been widely studied for their applicability to imaging rationale as well as outcomes.

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

A rigorous and validated process involving multiple societies and other stakeholders was used to develop the Appropriate Use Criteria (AUC). The AUC have been validated by various studies, including the ones cited earlier in this application. They are not merely expert panels but purposefully balanced committees undergoing a rigorous consensus process beyond even those used by guideline panels for decision making. A RAND modified Delphi process is used to determine the AUC rating that combines expert opinion with available evidence and specific patient information. The methods for this review

have been published and are available at: http://www.onlinejacc.org/content/71/8/935?_ga=2.169985062.746725178.1574208699-1575853885.1561572054 and https://www.acc.org/guidelines#tab4. Few studies are conducted to demonstrate a lack of benefit and thus, clinical risk and expert opinion is required to develop the AUC. Guidelines on the topic and references supporting recommendations related to the AUC clinical indications were identified. Additional literature searches were conducted to complete the available evidence published since the last guideline update. Specific evidence grades are not assigned by AUC, but generally diagnostic imaging evidence is based on observational studies, including well known risk models such as Framingham and Diamond and Forrester. In addition, a RAND modified Delphi process is used to determine the AUC rating that combines expert opinion with available evidence and specific patient information. Few studies are conducted to demonstrate a lack of benefit and thus, clinical risk and expert opinion is required to develop the AUC

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

Original

Douglas PS, Khandheria B, Stainback RF, ACCF/ASE/ACEP/AHA/ASNC/SCAI/SCCT/SCMR2008 appropriateness criteria for stress echocardiography. J Am Coll Cardiol. 2008;51:1127–47.

Hendel RH, Berman DS, Di Carli MF, et al. ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 Appropriate Use Criteria for Cardiac Radionuclide Imaging. J Am Coll Cardiol. 2009;53:2201–29.

Hendel RC, Patel MR, Kramer CM, Poon M. ACCF/ACR/SCCT/SCMR/ASNC/NASCI/SCAI/SIR 2006 appropriateness criteria for cardiac computed tomography and cardiac magnetic resonance imaging. J Am Coll Cardiol 2006;48:1475–97.

Updated

Wolk MJ, Bailey SR, Doherty JU et al. ACCF/AHA/ASE/ASNC/HFSA/HRS/SCAI/SCCT/SCMR/STS 2013 multimodality appropriate use criteria for the detection and risk assessment of stable ischemic heart disease. J Am Coll Cardiol 2014;63:XXX–XX.

The Appropriate Use Criteria have been published and updated on a regular basis by the American College of Cardiology in partnership with other societies and stakeholders. The evidence underlying the AUC appear in guidelines and systematic reviews contained in the appendix materials for these documents. The clinical indications and expert opinion used have been widely studied for their applicability to imaging rationale as well as outcomes.

## 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.*

Appropriate use criteria define “when to do” and “how often to do” a given procedure in the context of scientific evidence, the health care environment, the patient’s profile and a physician’s judgment. While practice guidelines provide a foundation for summarizing evidence-based cardiovascular care or for providing expert consensus opinions, in many areas, marked variability remains in the use of cardiovascular procedures, raising questions about over-use and under-use. Appropriate use criteria provide practical tools to measure this variability and to look at utilization patterns. The criteria are designed to examine the use of diagnostic and therapeutic procedures to support efficient use of medical resources, while also providing patients with quality, appropriate care.

A measure that reports rates of inappropriate imaging within practices would contain information regarding both cost and quality, because an inappropriate test results in both higher costs and poorer-quality care. Conversely, a reduction in this rate would simultaneously improve quality and decrease cost. Improvements in this metric should lead to consistent application of AUC and improve the efficiency of the system.

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

Hendel RC, Cerqueira M, Douglas PS. A multicenter assessment of the use of single-photon emission computed tomography myocardial perfusion imaging with appropriateness criteria. J Am Coll Cardiol. 2010 Jan 12;55(2):156-62.

These site specific performance scores were provided by a sub-analysis of the data collected for the above study.

Six sites participated in this pilot study; 3 urban, 2 suburban, and 1 rural location. Practices were located in Florida, Wisconsin, Oregon, and Arizona, and the number of cardiologists at each site ranged from 7 to 20 physicians. The number of SPECT MPI patients submitted from each site varied from 328 to 1,597 patients. A total of 6,351 subjects with complete

data were entered into the pilot database.

All Sites - 3.4%

Site 1 - 2.5%

Site 2 - 4.8%

Site 3 - 2.8%

Site 4 - 0.9%

**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.**

Hendel RC, Cerqueira M, Douglas PS. A multicenter assessment of the use of single-photon emission computed tomography myocardial perfusion imaging with appropriateness criteria. J Am Coll Cardiol. 2010 Jan 12;55(2):156-62.

See above

Fonseca R, Negishi K, Otahal P, et al. Temporal Changes in Appropriateness of Cardiac Imaging. J Am Coll Cardiol. 2015 Mar 3;65(8):763-73.

Mehta R, Agarwal S, Chandra S, Ward RP, Williams KA: Evaluation of the American College of Cardiology Foundation/American Society of Nuclear Cardiology appropriateness criteria for SPECT myocardial perfusion imaging. J Nucl Cardiol. 2008;5:337–44.

There were 1,623 patients (mean age 61 years ± 11, 61% males). Most common indications for SPECT were evaluation of ischemic equivalent for coronary artery disease (CAD), risk assessment post-revascularization, and preoperative evaluation for non-cardiac surgery. 10% of referrals were classified as inappropriate, 5% uncertain, and 3% unclassified. Appropriate referrals had a higher proportion of abnormal SPECT results than inappropriate referrals (40% vs 27%, OR 2.08, 95% CI 1.56-2.77, P < .001).

Ward RP, Al-Mallah MH, Grossman GB, Hansen CL, Hendel RC, Kerwin TC, McCallister BD Jr., Mehta R, Dm Polk, Tilkemeier PL,Vashist A, Williams KA, Wolinsky DG, Ficaro EP: American Society of Nuclear Cardiology: American Society of Nuclear Cardiology review of the ACCF/ASNC appropriateness criteria for single-photon emission computed tomography myocardial perfusion imaging

SPECT MPI). J Nucl Cardiol. 2007;14:e26–38.

Gibbons RJ, Miller TD, Hodge D, Urban L, Araoz PA, Pellikka P, McCully RB: Application of appropriateness criteria to stress single photon emission computed tomography sestamibi studies and stress echocardiograms in an academic medical center. J Am Coll Cardiol. 2008;51:1283–9.

The purpose of this study was to apply published appropriateness criteria for single-photon emission computed tomography (SPECT) myocardial perfusion imaging (MPI) in a single academic medical center.

The study retrospectively examined 284 patients who underwent stress SPECT MPI and 298 patients who underwent stress echocardiography before publication of the criteria.

5% of imaging was performed in asymptomatic patients within two years of prior PCI.

Carryer DJ, Hodge DO, Miller TD, Askew JW, Gibbons RJ. Application of appropriateness criteria to stress single photon emission computed tomography sestamibi studies: a comparison of the 2009 revised appropriateness criteria to the 2005 original criteria. Am Heart J. 2010 Aug;160(2):244-9..

An equal percentage of inappropriate studies (10.3%) were due to pre-op evaluation for low to intermediate risk non-cardiac surgery (indications 40, 41, and 42) and asymptomatic patients, less than 2 years after PCI (indication no. 59).

Bagai A, Eberg M, Koh M, Cheema AN, Yan AT, Dhoot A, Bhavnani SP, Wijeysundera HC, Bhatia RS, Kaul P, Goodman SG, Ko DT. Population-Based Study on Patterns of Cardiac Stress Testing After Percutaneous Coronary Intervention. Circ Cardiovasc Qual Outcomes. 2017 Oct;10(10).

**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.*

None

**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**

Mudrick DW, Shah BR, McCoy LA, at al. Patterns of stress testing and diagnostic catheterization after coronary stenting in 250 350 medicare beneficiaries. Circ Cardiovasc Imaging. 2013 Jan 1;6(1):11-9.

Several clinical risk factors at time of index PCI were associated with decreased likelihood of downstream testing (ST or CA, P<0.05 for all), including older age (hazard ratio [HR] 0.784 per 10-year increase), male sex (HR 0.946), heart failure (HR 0.925), diabetes mellitus (HR 0.954), smoking , (HR 0.804), and renal failure (HR 0.880).

Peterson T, Askew JW, Bell M, et al. Low yield of stress imaging in a population-based study of asymptomatic patients after percutaneous coronary intervention. Circ Cardiovasc Imaging. 2014 May;7(3):438-45.

Compared with patients who were asymptomatic at the time of stress imaging, patients who did not undergo any followup procedures (stress imaging, angiography, or coronary artery bypass grafting) after the index PCI were older, were more likely to have comorbidities

Luca SR, Koh M, Qiu F, Alter DA, Bagai A, Bhatia RS, Czarnecki A, Goodman SG, Lau C, Wijeysundera HC, Ko DT. Stress testing after percutaneous coronary interventions: a population-based study. CMAJ Open. 2017 May 26;5(2):E417-E423

The authors found that stress tests were not performed in accordance with patients´ higher baseline risk of adverse outcomes or risk of restenosis. Instead, many nonclinical factors, such as patients´ socioeconomic status and hospitals´ teaching status, were associated with higher use of stress tests.

## 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):*

Cardiovascular, Cardiovascular : Coronary Artery Disease (PCI)

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

Safety : Overuse

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

Elderly, Populations at Risk

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

No current webpage; only NQF specifications page

**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:** Imaging-Efficiency-Measures-Micro-specifications_Measure_Maintenance-635231485653419342.doc

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

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

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

Not an instrument-based measure

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

No

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

No changes have been made since endorsement.

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

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

Number of stress SPECT MPI, stress echo, CCTA and CMR performed in asymptomatic patients within 2 years of the most recent PCI

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

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

For all orders post PCI, determine all orders that were in asymptomatic patients:

Among asymptomatic patients, subtract date of most recent PCI from date of test requisition and categorize into orders less than two years since most recent PCI and orders placed greater than or equal to two years since most recent PCI

Patients qualify for this measure if:

- Asymptomatic AND
- Less than two years since most recent PCI

NOTE: Data collection from patient requisition is required to adequately determine patient's symptom status. Determination with only administrative data is not possible for these measures.

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

Number of stress SPECT MPI, stress echo, CCTA and CMR performed

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

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

All consecutive stress SPECT MPI, stress echocardiography, CCTA and CMR orders

Measurement Entity: Imaging laboratory prospectively measured on test requisition forms and/or patient charts

Level of Measurement/Analysis: Imaging laboratory*

*Attribution for inappropriate use is shared between the ordering physician and imaging laboratory. In an ideal world, attribution to the ordering physician or institution, as well as the imaging laboratory, would be reflected in the reporting of these measures. However, there are numerous complexities that prevent

assignment of these measures to individual ordering physicians. For example, ordering volumes from individual physicians and institutions are insufficient to make meaningful comparisons to allow such attribution. Thus, these measures will be reported at the level of the imaging laboratory. However, the extent to which the institution housing the imaging laboratory can impact these measures will be dependent upon cooperation of ordering physicians with the imaging laboratory.

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

None

**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.)*

None

**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.)*

None

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

No risk adjustment or risk stratification

If other:

**S.12. Type of score:**

Rate/proportion

If other:

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

Better quality = 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.)*

Locate all stress SPECT MPI, stress echocardiography, CCTA and CMR orders performed during the sampling period.

Record the total number of tests during the sampling period as the denominator.

From this sets of test orders, identify orders containing the criteria listed in the numerator

**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.

Measures are to be developed based on a sample of a full calendar year based on the following sampling methodology:

Select a starting month:

o January

- March
- May
- July
- September
- November

Begin 60 day data collection period on the 1st on the month for the selected starting month

Determine whether at least 30 stress SPECT, stress echo, CCTA and CMR orders have been placed during the selected time period. If not, select another time period with a minimum number of 30 cases. If no time period includes the minimum number of cases, then the imaging laboratory does not have sufficient volume to report this measure.

Sampling is required for this measure as full year data collection does not alter performance rates for this measure and would place an additional data collection burden on laboratories. It also allows laboratories to share performance with ordering physicians more quickly than would be possible under full year calendar reporting.

**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.

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

*If other, please describe in S.18.*

Other, Registry Data

**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.

Optimization of Patient Selection for Cardiac Imaging

**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)*

Available in attached appendix at A.1

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

Clinician : Group/Practice, Facility

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

Outpatient Services

If other:

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

**2. Validity – See attached Measure Testing Submission Form**

nqf_testing_attachment_7.1_671_July_2018_updated.docx

**2.1 For maintenance of endorsement**

*Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment.*

*Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

No

**2.2 For maintenance of endorsement**

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

No

**2.3 For maintenance of endorsement**

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

No - This measure is not risk-adjusted

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

**Measure Number** (*if previously endorsed*)**:** 671
**Measure Title**: Cardiac stress imaging not meeting appropriate use criteria: Routine testing after percutaneous coronary intervention (PCI)
**Date of Submission**: 11/1/2019

**Type of Measure:**

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

**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: Click here to describe | ☐ other: Click here to describe |

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

N/A – see below

**1.3. What are the dates of the data used in testing**? August 15, 2007 and May 15, 2010

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

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

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

11 practices encompassing 12 ZIP codes within the Chicago metropolitan area; 20 primary care physicians and 2 cardiologists.

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging. Circulation. 2013 Oct 8;128(15):1634-43. doi: 10.1161/CIRCULATIONAHA.113.002744. Epub 2013 Sep 10.

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

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Circulation. 2013 Oct 8;128(15):1634-43. doi: 10.1161/CIRCULATIONAHA.113.002744. Epub 2013 Sep 10

Baseline Clinical and Imaging Characteristics

| | Overall Cohort (n=1511) |
|---|---|
| Age, y | 59±13 |
| Women, n (%) | 657 (43.5) |
| Primary indication for MPI, n (%) | |
| Chest pain | 688 (45.5) |
| Dyspnea | 158 (10.5) |
| Abnormal ECG | 136 (9.0) |
| Evaluation of known CAD | 159 (10.5) |
| Preoperative assessment | 37 (2.4) |
| Syncope | 21 (1.4) |
| Asymptomatic | 262 (17.3) |
| Hypertension, n (%) | 841 (55.6) |
| Diabetes mellitus, n (%) | 333 (22.0) |
| Dyslipidemia, n (%) | 695 (46.0) |
| Tobacco use, n (%) | 181 (12.0) |
| Family history of CAD, n (%) | 544 (36.0) |
| Framingham 10-y CHD risk, % | 13±10 |
| Likelihood of obstructive CAD, %* | 18±13 |
| Exercise stress (Bruce) protocol, n (%) | 1164 (77.0) |
| BMI, kg/m$^2$ | 30±5.7 |
| Known CAD, n (%) | 271 (17.9) |
| Previous CABG, n (%) | 76 (5.0) |
| Previous PCI, n (%) | 87 (5.8) |
| Previous MI, n (%) | 37 (2.4) |
| Statin, n (%) | 580 (38.4) |

| | Overall Cohort (n=1511) |
|---|---|
| Antiplatelet, n (%) | 370 (24.5) |
| β-Blocker, n (%) | 307 (20.3) |
| ACE-I or ARB, n (%) | 567 (37.5) |
| Myocardial perfusion, n (%) | |
| Normal (SSS=0–3) | 1344 (88.9) |
| Mildly abnormal (SSS=4–8) | 79 (5.2) |
| Moderately abnormal (SSS=9–13) | 47 (3.1) |
| Severely abnormal (SSS >13) | 41 (2.7) |
| Myocardial ischemia, n (%) | |
| None (SDS ≤1) | 1399 (92.6) |
| Mild (SDS=2–4) | 38 (2.5) |
| Moderate (SDS=5–7) | 40 (2.6) |
| Severe (SDS >7) | 43 (2.8) |
| Type of perfusion abnormality, n (%) | |
| Reversible | 87 (5.8) |
| Fixed | 61(4.0) |
| Reversible and fixed | 19 (1.3) |
| Poststress LVEF <50%, n (%) | 78 (5.2) |

**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 cohort used for the validity testing is described above. A smaller single center study was used for reliability testing and is cited below.

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging.

The demographics of the single center study are as follows: May 1, 2005, and May 15, 2005. Mayo Clinic (Rochester, Minn). The mean±SD age of the 298 study patients was 66±13 years; 52% were men, 20% had diabetes mellitus, 60% had hypertension, 66% had hyperlipidemia, 54% had a history of smoking, 11% had a prior myocardial infarction, 20% had prior coronary revascularization, 36% had chest pain, 38% had dyspnea, and 41% had a normal resting ECG.

**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.

N/A

---

**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*)

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging.

Using the appropriateness criteria document, 2 experienced cardiac registered nurse abstractors reviewed patient demographics and other relevant information and classified each patient as appropriate, inappropriate, or uncertain. Patients who did not fit any of the clinical situations in the appropriateness criteria were judged to be not classifiable. The level of agreement between the 2 raters was analyzed.

Also, see section 2b1 for validity testing of data elements.

**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*)

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging. 2009 May;2(3):213-8.

Nurse abstract0r agreement kappa=0.72 for stress echocardiography

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

The data elements required for calculation of the appropriate use metrics can be obtained reliably by clinical staff from data residing in patient records with a high degree of agreement between nurses who would enter the data into the registry/clinical database.

---

**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)*
Relationship between appropriate use score and predictive value of SPECT MPI.

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

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Patients in the appropriate/uncertain group experienced significantly higher overall rates of death (HR, 2.9; 95% CI, 1.05–8.0; P=0.04), the composite of death or MI (HR, 1.04; 95% CI, 1.01–1.07; P=0.03), and the composite of cardiac death or MI (HR=5.7; 95% CI, 1.3–25.6; P=0.02) after adjustment for clinical covariates. Among patients in the appropriate/uncertain group, abnormal MPI continued to predict a multifold increase in the risk of death, cardiac death, composite of death or MI, and composite of cardiac death or MI (Figure 4). However, in the inappropriate group, there were no statistically significant differences in MACE rates between subjects with abnormal versus normal MPI (Figure 4). Furthermore, using Cox regression models, no interaction was identified between the study group and MPI finding in predicting death, the composite of death or MI, or the composite of cardiac death or MI (P=0.91, 0.70, and 0.43, respectively).

A Cox regression model demonstrated that inappropriate MPI use was a negative predictor of all-cause mortality (HR, 0.26; 95% CI, 0.10–0.67; P=0.005) after adjustment for myocardial perfusion finding (normal versus abnormal; HR, 2.5; 95% CI, 1.1–5.9; P=0.04) and depressed LVEF (<50%; HR, 3.7; 95% CI, 1.5–9.3; P=0.006); undergoing early coronary revascularization was not predictive of mortality (P=0.98). Similarly, in separate models, we demonstrated that inappropriate use was an independent negative predictor of the secondary end points of death or MI (HR, 0.31; 95% CI, 0.14–

0.70; P=0.005) and cardiac death or MI (HR, 0.16; 95% CI, 0.04–0.71; P=0.02) after adjustment for depressed LVEF, myocardial perfusion findings, and early revascularization. In these models, MPI and depressed LVEF independently predicted the composite end points of death or MI and cardiac death or MI, whereas undergoing early coronary revascularization after MPI was not predictive of these end points (P≥0.97). Finally, in forward stepwise Cox regression models, appropriate use was shown to have incremental prognostic value to perfusion imaging and depressed LVEF in predicting MACE; undergoing early revascularization (<60 days) did not provide significant additional predictive value

**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?*)
Appropriateness of imaging as measured by these metrics is correlated with the downstream value of the test in contributing to clinical decision making. As such, the metrics contribute to ensuring the prognostic value of the imaging procedures measured.

_____

**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*)

**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*)

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

_____

**2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES**
***If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.***

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **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.**

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

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

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

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

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

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

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

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

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

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

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

**2b3.9. Results of Risk Stratification Analysis**:

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

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

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

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

To date, there has been consistency in the studies showing the gap in performance on these metrics across sites. While individual practitioner level measurement and rates based on type of physician have shown variability, practice/hospital performance has been similar at baseline and after intervention to improve.

**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*)

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

No statistical tests have been applied to demonstrate differences among the measured entities at the practice/hospital level. Inappropriate use rates for individual practitioners ranged from 10% to 77% (P<0.001) and were higher among primary care physicians than cardiologists (47% versus 28%; P<0.001)

Fonseca R, Negishi K, Otahal P, et al. Temporal Changes in Appropriateness of Cardiac Imaging. J Am Coll Cardiol. 2015 Mar 3;65(8):763-73.

A separate meta-analysis demonstrated wide variation of appropriate use rates as described in the performance scores over time in section 2.b.1.

**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?*)

Inappropriate use is common among a wide range of practices and hospitals. Variability exists within practices and provides opportunities for peer to peer learning and improvement on these measures, especially within a practice or between primary care and specialists.

**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*)

**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*)

**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*)

---

**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*)

All subjects were classifiable according to the 2009 AUC and therefore no analysis for missing data was required. For validity testing, some subjects were lost to follow-up. Their demographic and AUC patterns were analyzed for similarity to the included patient cohort.

**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*)

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Compared with subjects with complete follow-up, the patients excluded (n=182) or lost to follow-up (n=14) were younger (mean age, 55±15 versus 59±13 years; P=0.001) and had lower likelihood of obstructive CAD (15±13% versus 18±13%; P=0.007) but similar mean 10-year Framingham coronary heart disease risk (12.7±10.8% versus 12.8±10%; P=0.88) and CAD prevalence (19% versus 18%; P=0.62). The prevalence of depressed LVEF and abnormal perfusion was nearly identical (P=0.97 and 0.89, respectively), with a similar breakdown of reversible, fixed, and mixed defects (P=0.64). The

excluded patients had a similar distribution of AUC classifications: 104 (53.1%) appropriate, 89 (45.4%) inappropriate, and 3 (1.5%) uncertain (P=0.53).

**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*)

Performance results were not biased as no missing data was recorded for the metrics themselves. Validity testing showed similar distribution of AUC, perfusion defects, and CAD prevalence and thus unlikely to have impacted results of this testing.

## 3. Feasibility

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

**3a. Byproduct of Care Processes**

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

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

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Other

If other: An EHR or Web portal prompts for clinical information in a decision support tool for individual cases that then are transmitted to a measurement registry

**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**.

Some data elements are in defined fields in electronic sources

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

Some data elements should already be a part of the electronic record (PCI history, scheduled surgery). In addition, e-ordering for diagnostic testing has been proposed for meaningful use, encouraging integration of these types of data elements. In addition, ACC has developed clinical decision support tools that can be embedded in electronic health records to capture the necessary information.

**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.**

Hendel, RC; Cerqueira, M; Douglas, PS et al. "A Multicenter Assessment of the Use of Single-Photon Emission Computed Tomography Myocardial Perfusion Imaging With Appropriateness Criteria". J Am Coll Cardiol. Published online December 10, 2009.

This study demonstrated the feasibility of data collection as well as the most frequent inappropriate indications. This allowed ACC to narrow the number of indications measured for this measure set along with the associated data elements.

**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)*.

None required. Decision support tools are available to aid in data collection and are available on a per test basis.

## 4. Usability and Use

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

**4a. Accountability and Transparency**

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

**4.1. Current and Planned Use**

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

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|

| | |
|---|---|
| | Public Reporting<br>PQRS<br>http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html<br>Payment Program<br>QPP<br>https://qpp.cms.gov/mips/<br>QPP<br>https://qpp.cms.gov/mips/<br>Regulatory and Accreditation Programs<br>IAC<br>http://www.intersocietal.org/intersocietal.htm<br>Professional Certification or Recognition Program<br>FOCUS<br>www.cardiosource.org/focus<br>Quality Improvement (Internal to the specific organization)<br>FOCUS<br>https://www.acc.org/focus |

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

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

QPP/MIPS - CMS/pay for performance/national. The data collected at the lab level for this measure can be further segmented by physician to help them understand their appropriate use patterns; although small sample sizes can limit comparability for some providers.
FOCUS - ACC/lab accreditation, quality improvement and utilization management/national - 25,000 cases with concentrations in DE (100% for SPECT MPI) and Western PA (10% for SPECT MPI and stress echo for cardiologists) - addtional 6,000 cases
IAC - lab accreditation/national - 100% - 5% of lab tests performed on an annual basis

**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.**

through CMS

**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.**

through CMS

**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.**

through CMS

**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.**

Luca SR, Koh M, Qiu F, Alter DA, Bagai A, Bhatia RS, Czarnecki A, Goodman SG, Lau C, Wijeysundera HC, Ko DT. Stress testing after percutaneous coronary interventions: a population-based study. CMAJ Open. 2017 May 26;5(2):E417-E423

The authors observed a decrease in the use of stress testing after PCI procedures over time. Same study cited in "disparities".

**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.**

None have been identified at this time.

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

## 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.

No

**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.**

**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?**

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

**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.)**

## Appendix

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

Attachment **Attachment:** FOCUS_Data_Collection_Sheet.docx

## Contact Information

**Co.1 Measure Steward (Intellectual Property Owner):** American College of Cardiology

**Co.2 Point of Contact:** Amy, Dearborn, adearborn@acc.org, 202-375-6576-

**Co.3 Measure Developer if different from Measure Steward:** American College of Cardiology Foundation

**Co.4 Point of Contact:** Joseph, Allen, jallen@acc.org, 202-375-6463-

## 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.**

All individuals are volunteer members representing American College of Cardiology Foundation:

Pamela Douglas, MD, MACC

Joseph Allen, MA

Robert Hendel, MD, FACC

Joseph Cacchione, MD, FACC

Manuel Cerqueira, MD, FACC

Joseph Drozda, MD, FACC

Michael Picard, MD, FACC

Martha Radford, MD, FACC

Leslee Shaw, PhD, FACC

Allen Taylor, MD, FACC

Group developed list of proposed measures, specifications, definitions, justification, etc.

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 2009

**Ad.3 Month and Year of most recent revision:** 11, 2019

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

**Ad.5 When is the next scheduled review/update for this measure?** 12, 2020

**Ad.6 Copyright statement:** Copyright 2009. American College of Cardiology Foundation

**Ad.7 Disclaimers:**

**Ad.8 Additional Information/Comments:** dates above refer to maintenance of endorsement

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 #:** 0672

**Corresponding Measures:**

**De.2. Measure Title:** Cardiac stress imaging not meeting appropriate use criteria: Testing in asymptomatic, low risk patients

**Co.1.1. Measure Steward:** American College of Cardiology

**De.3. Brief Description of Measure:** Percentage of all stress SPECT MPI, stress echo, CCTA, and CMR performed in asymptomatic, low CHD risk patients for initial detection and risk assessment

**1b.1. Developer Rationale:** Appropriate use criteria define "when to do" and "how often to do" a given procedure in the context of scientific evidence, the health care environment, the patient's profile and a physician's judgment. While practice guidelines provide a foundation for summarizing evidence-based cardiovascular care or for providing expert consensus opinions, in many areas, marked variability remains in the use of cardiovascular procedures, raising questions about over-use and under-use. Appropriate use criteria provide practical tools to measure this variability and to look at utilization patterns. The criteria are designed to examine the use of diagnostic and therapeutic procedures to support efficient use of medical resources, while also providing patients with quality, appropriate care.

A measure that reports rates of inappropriate imaging within practices would contain information regarding both cost and quality, because an inappropriate test results in both higher costs and poorer-quality care. Conversely, a reduction in this rate would simultaneously improve quality and decrease cost. Improvements in this metric should lead to consistent application of AUC and improve the efficiency of the system.

**S.4. Numerator Statement:** Number of stress SPECT MPI, stress echo, CCTA, and CMR performed for asymptomatic, low CHD risk patients for initial detection and risk assessment*

**S.6. Denominator Statement:** Number of stress SPECT MPI, stress echo, CCTA, and CMR performed

**S.8. Denominator Exclusions:** None

**De.1. Measure Type:** Efficiency

**S.17. Data Source:** Other, Registry Data

**S.20. Level of Analysis:** Clinician : Group/Practice, Facility

**IF Endorsement Maintenance – Original Endorsement Date:** Apr 26, 2011 **Most Recent Endorsement Date:** Jun 29, 2015

## Preliminary Analysis: Maintenance of Endorsement

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

### Criteria 1: Importance to Measure and Report

#### 1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

**Summary of prior review in 2015**

- The developer provides evidence from the *2010 ACCF/AHA Guideline for Assessment of Cardiovascular Risk in Asymptomatic Adults*. The recommendations state that stress echocardiography and stress MPI are not indicated for cardiovascular risk assessment in low- or intermediate-risk asymptomatic adults; and, coronary computed tomography angiography and MRI for detection of vascular plaque are not recommended for cardiovascular risk assessment in asymptomatic adults.
    - Evidence is graded as "C", meaning "very limited patient populations evaluated. Only consensus opinion of experts, case studies, or standard of care." Recommendations are graded as "Class III: No Benefit," meaning "conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective, and in some cases may be harmful."
- The developer also includes a United States Preventive Services Task Force (USPSTF) recommendation against "screening with rest or exercise electrocardiography (ECG) for the prediction of coronary heart disease (CHD) in asymptomatic adults at low risk for CHD events."
    - The USPSTF gave the recommendation a "D" meaning "there is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits." USPSTF assigned the evidence a Grade I, meaning "the current evidence is insufficient to assess the balance of benefits and harms of the service."
- The developers noted a lack of studies in population-based asymptomatic individuals and therefore, they were unable to provide information on the quantity, quality and consistency of a body of evidence that the measured process leads to a desired health outcome.

**Changes to evidence from last review**

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

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

***Questions for the Committee:***

- How strong is the evidence for this relationship?
- Is the evidence directly applicable to the process of care being measured?

For possible exception to the evidence criterion:

- Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?
- Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?
- Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?

**Guidance from the Evidence Algorithm**

Process measure based on systematic review (Box 3) → Guidelines based on expert opinion (Box 7) → No empirical evidence (Box 10) → INSUFFICIENT

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

**RATIONALE:** The guidelines presented represent expert opinions in an area where few to no studies exist.

## 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.

- For maintenance measures, performance scores on the measure as specified at the specified level of analysis is required for maintenance of endorsement. The results provided do not appear to be calculated using the measure as specified. They do not include the same tests as the measure, and it is difficult to tell if the calculations were performed in alignment with the measure specifications.
- The developer presented site-specific performance score, which were obtained from a sub-analysis of the data collected for one study. The study is from 2010.
    - Six sites participated in the pilot study including 3 urban, 2 suburban, and 1 rural location in Florida, Wisconsin, Oregon, and Arizona. The number of cardiologists at each site ranged from 7 to 20 physicians. The number of SPECT MPI patients submitted from each site varied from 328 to 1,597 patients, but a total of 6,351 subjects with complete data were entered into the pilot database.
    - The developer provided results for four sites with results ranging from 3.5% to 8.8%. No specific information is provided about each of the site, i.e., size, number of studies, location, ownership, or the timeframe when the data were obtained.
    - There is not enough information to determine if the results provided correspond to the levels of analysis for which this measure is specified. The study only includes one of the four types of tests included in the measure.
- The developers provided data from the literature that indicated Appropriate referrals yield a higher proportion of abnormal SPECT results than inappropriate referrals (40% vs 27%, OR 2.08, 95% CI 1.56-2.77, P < .001).
- The developers provided a summary of another study that applied published appropriate use criteria (AUC) for single-photon emission computed tomography (SPECT) myocardial perfusion imaging (MPI) in a single academic medical center. The study retrospectively applied AUC to 284 patients who underwent stress SPECT MPI and 298 patients who underwent stress echocardiography and found that 48% of the inappropriate imaging was in low risk, asymptomatic patients.

**Disparities**

- Disparities data from the measure as specified is required for maintenance of endorsement. No disparities data or summary of disparities data from the literature

***Questions for the Committee:***

- Does the developer provide enough data to show a gap in care that warrants a national performance measure?
- Does the data provided demonstrate a need for this measure?
- Since the developer did not provide any information on disparities, are you aware of evidence that disparities exist in this area of healthcare?

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

**RATIONALE:** The data provided for performance gap and disparities is minimal or insufficient. The data provided are from 2010, providing no information on current performance gaps. Performance scores on the measure as specified are required for maintenance of endorsement. Those scores are not provided.

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

1a.

- N/A
- Old guideline based on level of evidence C
- Minimal studies. Evidence is mainly based on Bayesian statistics and that testing a low risk population will have higher false positives than true positives and a subsequent cascade of unnecessary expensive care could result without benefit to the patient or population
- Aligned with NQF assessment. Would be good to hear from the cardiologists on the committee.

1b.

- No performance gap data are presented. No disparities data either
- 2010 data from small sample
- no new data presented. No data on measure use presented
- Majority of data from a 2010 study. No gap analyzed from current data. Agree with NQF Staff

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability: Specifications and Testing**

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

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

### Reliability

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

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

## Validity

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

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

**Composite measures only:**

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

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

**Evaluators:** NQF Staff

Scientific Acceptability: Preliminary Analysis Form

**Measure Number:** 0672

**Measure Title:** Cardiac stress imaging not meeting appropriate use criteria: Testing in asymptomatic, low risk patients

### 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: Unclear**

### Measure is:

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

### RELIABILITY: SPECIFICATIONS

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

  **Submission document:** "MIF_0672" 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.*

- **Briefly summarize any concerns about the measure specifications.**
  - Does the measure include all ages? No age range is included in the specifications.
  - The developer indicates Clinician: Group/Practice is a level of analysis for this measure. It's unclear which clinician would be held accountable. The denominator of number of tests performed doesn't correspond to an ordering physician. Is it the performing physician? The attribution should be clear.

### RELIABILITY: TESTING

**Submission document:** "MIF_0672" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

- **Reliability testing level** ☐ **Measure score** ☒ **Data element** ☐ **Neither**
- **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
  ☐ **Yes** ☒ **No**
  - Information supplied in testing attachment does not appear to correspond to data source or levels of analysis indicated.

- 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**

- **Assess the method(s) used for reliability testing**

  **Submission document:** Testing attachment, section 2a2.2

  - The developer states reliability was tested at the data element level.
  - The study included to demonstrate reliability testing is a single-center study including 298 patients. It includes stress echocardiogram and SPECT MPI, but not the other cardiac tests included in the measure specifications.
  - The study included appears to focus on using appropriate use criteria to evaluate the appropriateness of a test whereas this measure attempts to identify tests performed on asymptomatic, low-risk patients. These are two different things.
  - The inter-rater reliability provided is for the level of agreement in two nurses' appropriateness ratings for the cardiac testing. Appropriateness ratings are not a data element of this measure. The relationship between the appropriateness ratings and the measure specifications is unclear.

- **Assess the results of reliability testing**

  **Submission document:** Testing attachment, section 2a2.3

  There is not enough information provided to assess the reliability of this measure or its data elements.

- 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)

  - Enough information not provided by the developer

- 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)

  - Enough information not provided by the developer

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

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

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

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

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

- **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
  - There is not enough information provided to assess the reliability of this measure or its data elements. The information provided in the reliability section is not clearly related to the measure score or to the data elements in the measure. Testing does not appear to correspond to the levels of analysis (clinician: group/practice and facility) indicated for the measure.
  - In addition to concerns with the testing, staff identified concerns with the clarity of the specifications, particularly clinician attribution.

### VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

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

  - The developer indicates that there are no exclusions for this measure.
  - The measure specifications indicate that patients with a previous CHD assessment by a list of methods, no matter the result, are not included in the measure. If an asymptomatic, low-risk patient had a previous inappropriate test, this seems to indicate they would not be included as asymptomatic and low-risk for future assessments.

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

  **Submission document:** Testing attachment, section 2b4.

  - The developer's discussion of differences in performance focuses on inappropriate use and it is unclear if these results are in line with the focus of this measure or a more general application of AUC.
  - The developer provides no details on statistical testing of measure results.
  - While the developer notes that there is variation in inappropriate use rates at the individual-practitioner level and that these rates vary by physician specialty, no method is highlighted to identify meaningful differences in performances. It is also unclear if these physician specialties are the groups to which the measure would be applied/attributed.

- Previously one of the committee members questioned whether the general data supplied proves reliability.

- **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
  **Submission document:** Testing attachment, section 2b5.
    - Not applicable

- **Please describe any concerns you have regarding missing data.**
  **Submission document:** Testing attachment, section 2b6.
  - The developer reports "All subjects were classifiable according to the 2009 AUC and therefore no analysis for missing data was required." It is unclear which patients are being referenced and the relationship between the data in the study and the data elements of this measure is unclear.
  - Previously, one committee member highlighted that the outcomes are inferred from only nuclear perfusion imaging and stress echocardiography.

- **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

### VALIDITY: TESTING

- **Validity testing level: ☐ Measure score ☐ Data element ☐ Both**
- **Method of establishing validity of the measure score:**
  - ☐ **Face validity**
  - ☐ **Empirical validity testing of the measure score**
  - ☒ **N/A (score-level testing not conducted)**
- **Assess the method(s) for establishing validity**

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

- The developer states their method of validity testing is the "relationship between appropriate use score and predictive value of SPECT MPI." This does not appear to align with empirical validity testing for measure 0672.
- Previously, the committee raised concerns that it is unclear what the sample size was and how the FOCUS questionnaire was implemented for each of the imaging modalities. This is relevant since the most expensive tests (CMR, NPI, and CT CA) are often under the direction of non-cardiology directors.

- **Assess the results(s) for establishing validity**
  - The results presented do not provide information that can be used to assess the validity of this measure. There is not enough relevant information provided.

  **Submission document: Testing attachment, section 2b2.3**

- **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)

- **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)

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

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

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

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

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

- **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**
  - The information provided in the validity section is not directly related to the measure score or to the data elements in the measure. There is not enough information provided to assess the validity of the measure score or the data elements.

## ADDITIONAL RECOMMENDATIONS

- **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.**

***Questions for the Committee regarding reliability:***

- Is it clear from the provided specifications how this measure would be attributed to a clinician group or practice and which clinician group or practice would be held accountable?
- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The staff was not satisfied with the reliability testing for the measure. Does the Committee agree with the staff assessment?

***Questions for the Committee regarding validity:***

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The staff was not satisfied with the validity analyses for the measure. Does the Committee agree with the staff assessment?

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

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

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

2a1.

- Much of the evidence is graded as insufficient despite the fact that this measure has been in use for many years.
- No mention of age limits
- I am not so worried about attribution to a given clinician which should not be done, this should be reported at a practice or higher level. Other data insufficeint to assess reliability
- Agree with NQF assessment of the reliability

2a2.

- Yes. It is graded as insufficient
- Based on single center study of ~300 patients
- Unlear as to how low risk/Asx are defined
- Agree with NQF assessment of the reliability

2b1.

- The measure is based on face validity—AUC
- No clear statistical testing provided of measure results
- Exclusion of previously tested pateints without known CHD seems inappropriate
- Agree with NQF assessment of the validity

2b4-7.

- The data have not been updated in the current application
- No clear statistical testing of measure results
- not a clean data entry vehicle
- Agree with NQF assessment of the validity

2b2-3

- No exclusions; no risk adjustment
- n/a

- No risk adjustment
- Agree with NQF assessment of the validity

## Criterion 3. Feasibility

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

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

- Data is generated by and used by healthcare personnel during the provision of care, (e.g., indication for testing), and is coded by someone other than person obtaining original information. Additionally, an EHR or Web portal prompts for clinical information in a decision support tool for individual cases that then are transmitted to a measurement registry
- Some data elements are in defined fields in electronic sources. And, ACC has developed clinical decision support tools that can be embedded in electronic health records to capture the necessary information.

***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?
- What is the burden of data collection, i.e., chart abstraction and data entry to a registry?

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

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

- The measure is feasible for participating entities
- no concerns
- Given mandated use of AUC, feasability is acceptable
- The data seem vague - either with how it is captured or obtained digitally. Not specific.

## Criterion 4: Usability and Use

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

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

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

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

**Current uses of the measure**

**Publicly reported?** ☒ **Yes** ☐ **No**

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

**Accountability program details**

- The developer states the measures is used in the following programs:
  - MIPS - CMS/pay for performance/national; The data collected at the lab level for this measure can be further segmented by physician to help them understand their appropriate use patterns; although small sample sizes can limit comparability for some providers
  - FOCUS - ACC/lab accreditation, quality improvement and utilization management/national - 25,000 cases with concentrations in DE (100% for SPECT MPI) and Western PA (10% for SPECT MPI and stress echo for cardiologists) - additional 6,000 cases
  - IAC - lab accreditation/national - this measure may be used in support of accreditation

**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**

- None.

**Additional Feedback:**

- None

***Questions for the Committee:***

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

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

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

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

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

**Improvement results**

- No results or improvement trends are provided.

**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 reports there have been no unexpected findings.

**Potential harms**

- According to the developers, no unintended consequences have been identified for this measure.

**Additional Feedback:**

***Questions for the Committee:***

- Are you aware of any unintended consequences for this measure?
- Do the benefits of the measure outweigh any potential unintended consequences?

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

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

4a1.

- The measure is in use by several quality programs
- Yes in MIPS
- Needs wider use, but appears to have use in at least two geographies
- Results are publicly reported.

4b1.

- No significant harms identified
- No concerns
- See above comment, no suspected harms
- Nothing listed/no harm found.

## Criterion 5: Related and Competing Measures

**Related or competing measures**

- 0671: Cardiac stress imaging not meeting appropriate use criteria: Routine testing after percutaneous coronary intervention (PCI)
- 0672: Cardiac stress imaging not meeting appropriate use criteria: Testing in asymptomatic, low risk patients

**Harmonization**

- None

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

- 0671 and 0672 are related but don't compete
- No concerns
- related measures all suffer from same data issues
- There seems to be a lot of overlap between 670-672 in both current material and scope. Would like to see how they augment each other

## Public and Member Comments

**Comments and Member Support/Non-Support Submitted as of: January 21, 2020**

- **No NQF members have submitted a support/non-support choice as of this date.**

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

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

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

NQF_evidence_attachment_Sep2017_-_672.docx

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

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

No.

### 1a. Evidence (subcriterion 1a)

**Measure Number** (*if previously endorsed*)**:** 672

**Measure Title**: Cardiac stress imaging not meeting appropriate use criteria: Testing in asymptomatic, low risk patients

**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**: 11/7/2019

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

Outcome

☐ Outcome: Resource Use and Avoidance of Negative Clinical Benefit Risk Ratio for Patient Population

☐ 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: Resource Use and Avoidance of Negative Clinical Benefit Risk Ratio for Patient Population

☒ Appropriate use measure: Cardiac stress imaging not meeting appropriate use criteria: Testing in asymptomatic, low risk patients

☐ 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.

Measurement of appropriate use summarizes the financial value/resources use and avoidance of a negative clinical benefit risk ratio across a patient population in which a procedure is used. Various factors influence the ability of a procedure to contribute to the diagnosis and treatment of a patient, including the clinical factors summarized by appropriate use measures. These clinical factors combined with physician and patient decision making determine the probability that a procedure will have the intended impact on health outcomes of the patient

**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.**

**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:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | **2010 ACCF/AHA Guideline for Assessment of Cardiovascular Risk in Asymptomatic Adults**<br>**A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines *Developed in Collaboration With the American Society of Echocardiography, American Society of Nuclear Cardiology, Society of Atherosclerosis Imaging and Prevention, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance***<br><br>**Philip Greenland, Joseph S. Alpert, George A. Beller, Emelia J. Benjamin, Matthew J. Budoff, Zahi A. Fayad, Elyse Foster, Mark A. Hlatky, John McB. Hodgson, Frederick G. Kushner, Michael S. Lauer, Leslee J. Shaw, Sidney C. Smith Jr, Allen J. Taylor, William S. Weintraub and Nanette K. Wenger**<br><br>**December 2010**<br><br>**J Am Coll Cardiol 2010;56:e50–103**<br><br>**https://www.sciencedirect.com/science/article/pii/S0735109710037186?via%3Dihub** |

| | |
|---|---|

| | |
|---|---|
| **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.** | **ACC/AHA guidelines**<br>**E76 – e81**<br>**CLASS III: NO BENEFIT**<br>**1. Stress echocardiography is not indicated for cardiovascular risk assessment in low- or intermediate-risk asymptomatic adults. (Exercise or pharmacologic stress echocardiography is primarily used for its role in advanced cardiac evaluation of symptoms suspected of representing CHD and/or estimation of prognosis in patients with known coronary artery disease or the assessment of patients with known or suspected valvular heart disease.) (Level of Evidence: C)**<br><br>**CLASS III: NO BENEFIT**<br>**1. Stress MPI is not indicated for cardiovascular risk assessment in low- or intermediate-risk asymptomatic adults (Exercise or pharmacologic stress MPI is primarily used and studied for its role in advanced cardiac evaluation of symptoms suspected of representing CHD and/or estimation of prognosis in patients with known CAD.) (326). (Level of Evidence: C)**<br><br>**CLASS III: NO BENEFIT**<br>**1. Coronary computed tomography angiography is not recommended for cardiovascular risk assessment in asymptomatic adults (372). (Level of Evidence: C)**<br><br>**CLASS III: NO BENEFIT**<br>**1. MRI for detection of vascular plaque is not recommended for cardiovascular risk assessment in asymptomatic adults. (Level of Evidence: C)** |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | Level of Evidence C: Very limited patient populations evaluated. Only consensus opinion of experts, case studies, or standard of care. |
| Provide all other grades and definitions from the evidence grading system | See below* |
| Grade assigned to the **recommendation** with definition of the grade | CLASS III: NO BENEFIT<br>Conditions for which there is evidence and/or general agreement that the procedure/treatment is |

| | |
|---|---|
| | not useful/effective, and in some cases may be harmful. |
| Provide all other grades and definitions from the recommendation grading system | See below* |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | Quantity: There are few studies on the role of stress MPI for risk assessment in asymptomatic persons. The guideline writing committee did not identify any studies in population-based (relatively unselected) asymptomatic individuals. Reported studies of stress perfusion imaging in asymptomatic persons have involved selected higher-risk patients who were referred for cardiac risk evaluation.<br><br>Quality: n/a |

| | |
|---|---|
| **Estimates of benefit and consistency across studies** | **This measure looks at the absence of potential benefit in a specific population which is derivative of the studies examined but not a direct end point of the studies reviewed.** |
| What harms were identified? | The studies did not examine harm. |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | N/a |


***** grades and definitions from the evidence grading system

| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | **Screening for coronary heart disease with electrocardiography: US Preventive Services Task Force recommendation statement.**<br><br>**Virginia A. Moyer, MD, MPH, on behalf of the U.S. Preventive Services Task Force**<br><br>**October 2012**<br><br>**Moyer VA, on behalf of the U.S. Preventive Services Task Force*. Screening for Coronary Heart Disease With Electrocardiography: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2012;157:512–518**<br><br>**https://annals.org/aim/fullarticle/1363528/screening-coronary-heart-disease-electrocardiography-u-s-preventive-services-task** |
| 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. | Page 513<br>The USPSTF recommends against screening with resting or exercise electrocardiography (ECG) for the prediction of coronary heart disease (CHD) events in asymptomatic adults at low risk for CHD events (D recommendation). |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | D – The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. |
| Provide all other grades and definitions from the evidence grading system | ** see below |
| Grade assigned to the **recommendation** with definition of the grade | Grade I (insufficient evidence). The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined. |
| Provide all other grades and definitions from the recommendation grading system | ** see below |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | Although evidence is insufficient to determine whether screening adults at increased risk is beneficial, those who are at intermediate risk for CHD events have the greatest potential for net benefit from ECG screening. Reclassification into a higher risk category might lead to more intensive medical management that could lower |

| | |
|---|---|
| | the risk for CHD events, but it might also result in harms, including such adverse medication effects as gastrointestinal bleeding and hepatic injury. The risk–benefit tradeoff would be most favorable if persons could be accurately reclassified from intermediate to high risk.<br><br>For asymptomatic adults at low risk for CHD events, the incremental information offered by resting or exercise ECG (beyond that obtained with conventional CHD risk factors) is highly unlikely to result in a change in risk stratification that would prompt interventions and ultimately reduce CHD-related events. |
| Estimates of benefit and consistency across studies | n/a |
| What harms were identified? | n/a |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | n/a |

**** GRADES AND DEFINITIONS, USPSTF**

| Grade | Definition | Suggestions for Practice |
|---|---|---|
| A | The USPSTF recommends the service. There is high certainty that the net benefit is substantial. | Offer or provide this service. |
| B | The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. | Offer or provide this service. |
| C | The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small. | Offer or provide this service for selected patients depending on individual circumstances. |
| D | The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. | Discourage the use of this service. |
| I Statement | The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined. | Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms. |

***Levels of Certainty Regarding Net Benefit***

| Level of Certainty* | Description |
|---|---|
| **High** | The available evidence usually includes consistent results from well-designed, well-conducted studies in representative primary care populations. These studies assess the effects of the preventive service on health outcomes. This conclusion is therefore unlikely to be strongly affected by the results of future studies. |
| **Moderate** | The available evidence is sufficient to determine the effects of the preventive service on health outcomes, but confidence in the estimate is constrained by such factors as:<br>• The number, size, or quality of individual studies.<br>• Inconsistency of findings across individual studies.<br>• Limited generalizability of findings to routine primary care practice.<br>• Lack of coherence in the chain of evidence.<br>As more information becomes available, the magnitude or direction of the observed effect could change, and this change may be large enough to alter the conclusion. |
| **Low** | The available evidence is insufficient to assess effects on health outcomes. Evidence is insufficient because of:<br>• The limited number or size of studies.<br>• Important flaws in study design or methods.<br>• Inconsistency of findings across individual studies.<br>• Gaps in the chain of evidence.<br>• Findings not generalizable to routine primary care practice.<br>• Lack of information on important health outcomes.<br>More information may allow estimation of effects on health outcomes. |

*The USPSTF defines certainty as "likelihood that the USPSTF assessment of the net benefit of a preventive service is correct." The net benefit is defined as benefit minus harm of the preventive service as implemented in a general, primary care population. The USPSTF assigns a certainty level based on the nature of the overall evidence available to assess the net benefit of a preventive service.

---

**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.*

Measurement of appropriate use summarizes the financial value/resources use and avoidance of a negative clinical benefit risk ratio across a patient population in which a procedure is used. Various factors influence the ability of a procedure to contribute to the diagnosis and treatment of a patient, including the clinical factors summarized by appropriate use measures. These clinical factors combined with physician and patient decision making determine the probability that a procedure will have the intended impact on health outcomes of the patient.

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Circulation. 2013 Oct 8;128(15):1634-43. doi: 10.1161/CIRCULATIONAHA.113.002744. Epub 2013 Sep 10.

Patients in the appropriate/uncertain group experienced significantly higher overall rates of death (HR, 2.9; 95% CI, 1.05–8.0; P=0.04), the composite of death or MI (HR, 1.04; 95% CI, 1.01–1.07; P=0.03), and the composite of cardiac death or MI (HR=5.7; 95% CI, 1.3–25.6; P=0.02) after adjustment for clinical covariates. Among patients in the appropriate/uncertain group, abnormal MPI continued to predict a multifold increase in the risk of death, cardiac death, composite of death or MI, and composite of cardiac death or MI (Figure 4). However, in the inappropriate group, there were no statistically significant differences in MACE rates between subjects with abnormal versus normal MPI (Figure 4). Furthermore, using Cox regression models, no

interaction was identified between the study group and MPI finding in predicting death, the composite of death or MI, or the composite of cardiac death or MI (P=0.91, 0.70, and 0.43, respectively).

A Cox regression model demonstrated that inappropriate MPI use was a negative predictor of all-cause mortality (HR, 0.26; 95% CI, 0.10–0.67; P=0.005) after adjustment for myocardial perfusion finding (normal versus abnormal; HR, 2.5; 95% CI, 1.1–5.9; P=0.04) and depressed LVEF (<50%; HR, 3.7; 95% CI, 1.5–9.3; P=0.006); undergoing early coronary revascularization was not predictive of mortality (P=0.98). Similarly, in separate models, we demonstrated that inappropriate use was an independent negative predictor of the secondary end points of death or MI (HR, 0.31; 95% CI, 0.14–0.70; P=0.005) and cardiac death or MI (HR, 0.16; 95% CI, 0.04–0.71; P=0.02) after adjustment for depressed LVEF, myocardial perfusion findings, and early revascularization. In these models, MPI and depressed LVEF independently predicted the composite end points of death or MI and cardiac death or MI, whereas undergoing early coronary revascularization after MPI was not predictive of these end points (P≥0.97). Finally, in forward stepwise Cox regression models, appropriate use was shown to have incremental prognostic value to perfusion imaging and depressed LVEF in predicting MACE; undergoing early revascularization (<60 days) did not provide significant additional predictive value

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

Diagnostic testing, such as stress SPECT MPI, stress echocardiography, CCTA, and CMR, is used to detect disease and provide risk assessment used to modify treatment strategies and approaches. Information provided by such testing can initiate, modify and stop further treatments for coronary heart disease (medications and revascularization) which have an impact on patient outcomes. In addition, false positives and false negatives can adversely impact the patient and their treatment outcomes. Lastly, radiation from stress SPECT MPI and CTA poses a minimal but still important consideration for patient safety. Ensuring proper patient selection can avoid using resources in patients not expected to benefit from the testing and for which the associated risks would be unnecessary.

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

A rigorous and validated process involving multiple societies and other stakeholders was used to develop the Appropriate Use Criteria (AUC). The AUC have been validated by various studies, including the ones cited earlier in this application. They are not merely expert panels but purposefully balanced committees undergoing a rigorous consensus process beyond even those used by guideline panels for decision making. A RAND modified Delphi process is used to determine the AUC rating that combines expert opinion with available evidence and specific patient information. The methods for this review have been published and are available at: http://www.onlinejacc.org/content/71/8/935?_ga=2.169985062.746725178.1574208699-1575853885.1561572054 and https://www.acc.org/guidelines#tab4. Few studies are conducted to demonstrate a lack of benefit and thus, clinical risk and expert opinion is required to develop the AUC. Guidelines on the topic and references supporting recommendations related to the AUC clinical indications were identified. Additional literature searches were conducted to complete the available evidence published since the last guideline update. Specific evidence grades are not assigned by AUC, but generally diagnostic imaging evidence is based on observational studies, including well known risk models such as Framingham and Diamond and Forrester. In addition, a RAND modified Delphi process is used to determine the AUC rating that combines expert opinion with available evidence and specific patient information. Few studies are conducted to demonstrate a lack of benefit and thus, clinical risk and expert opinion is required to develop the AUC

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

Each of the documents below covers a clinical imaging procedure and was developed using the AUC methodology cited below.

The Appropriate Use Criteria have been published and updated on a regular basis by the American College of Cardiology in partnership with other societies and stakeholders. The evidence underlying the AUC appear in guidelines and systematic reviews contained in the appendix materials for these documents. The clinical indications and expert opinion used have been widely studied for their applicability to imaging rationale as well as outcomes.

Original

Douglas PS, Khandheria B, Stainback RF, ACCF/ASE/ACEP/AHA/ASNC/SCAI/SCCT/SCMR2008 appropriateness criteria for stress echocardiography. J Am Coll Cardiol. 2008;51:1127–47.

Hendel RH, Berman DS, Di Carli MF, et al. ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 Appropriate Use Criteria for Cardiac Radionuclide Imaging. J Am Coll Cardiol. 2009;53:2201–29.

Hendel RC, Patel MR, Kramer CM, Poon M. ACCF/ACR/SCCT/SCMR/ASNC/NASCI/SCAI/SIR 2006 appropriateness criteria for cardiac computed tomography and cardiac magnetic resonance imaging. J Am Coll Cardiol 2006;48:1475–97.

Updated

Wolk MJ, Bailey SR, Doherty JU et al. ACCF/AHA/ASE/ASNC/HFSA/HRS/SCAI/SCCT/SCMR/STS 2013 multimodality appropriate use criteria for the detection and risk assessment of stable ischemic heart disease. J Am Coll Cardiol 2014;63:XXX–XX.

The Appropriate Use Criteria have been published and updated on a regular basis by the American College of Cardiology in partnership with other societies and stakeholders. The evidence underlying the AUC appear in guidelines and systematic reviews contained in the appendix materials for these documents. The clinical indications and expert opinion used have been widely studied for their applicability to imaging rationale as well as outcomes.

## 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.*

Appropriate use criteria define "when to do" and "how often to do" a given procedure in the context of scientific evidence, the health care environment, the patient's profile and a physician's judgment. While

practice guidelines provide a foundation for summarizing evidence-based cardiovascular care or for providing expert consensus opinions, in many areas, marked variability remains in the use of cardiovascular procedures, raising questions about over-use and under-use. Appropriate use criteria provide practical tools to measure this variability and to look at utilization patterns. The criteria are designed to examine the use of diagnostic and therapeutic procedures to support efficient use of medical resources, while also providing patients with quality, appropriate care.

A measure that reports rates of inappropriate imaging within practices would contain information regarding both cost and quality, because an inappropriate test results in both higher costs and poorer-quality care. Conversely, a reduction in this rate would simultaneously improve quality and decrease cost. Improvements in this metric should lead to consistent application of AUC and improve the efficiency of the system.

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

Hendel RC, Cerqueira M, Douglas PS. A multicenter assessment of the use of single-photon emission computed tomography myocardial perfusion imaging with appropriateness criteria. J Am Coll Cardiol. 2010 Jan 12;55(2):156-62.

These site specific performance scores were provided by a sub-analysis of the data collected for the above study.

Six sites participated in this pilot study; 3 urban, 2 suburban, and 1 rural location. Practices were located in Florida, Wisconsin, Oregon, and Arizona, and the number of cardiologists at each site ranged from 7 to 20 physicians. The number of SPECT MPI patients submitted from each site varied from 328 to 1,597 patients. A total of 6,351 subjects with complete

data were entered into the pilot database.

All Sites - 6.3%

Site 1 - 6.8%

Site 2 - 8.8%

Site 3 - 5.7%

Site 4 - 3.5%

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

Hendel RC, Cerqueira M, Douglas PS. J Am Coll Cardiol. 2010 Jan 12;55(2):156-62. doi: 10.1016/j.jacc.2009.11.004. A multicenter assessment of the use of single-photon emission computed tomography myocardial perfusion imaging with appropriateness criteria.

See above

Fonseca R, Negishi K, Otahal P, et al. Temporal Changes in Appropriateness of Cardiac Imaging. J Am Coll Cardiol. 2015 Mar 3;65(8):763-73.

Krumholz HM, Keenan PS, Brush JE et al. Standards for measures used for public reporting of efficiency in health care. J Am Coll Cardiol. 2008 Oct 28;52(18):1518-26.

Mehta R, Agarwal S, Chandra S, Ward RP, Williams KA: Evaluation of the American College of Cardiology Foundation/American Society of Nuclear Cardiology appropriateness criteria for SPECT myocardial perfusion imaging. J Nucl Cardiol. 2008;5:337–44.

There were 1,623 patients (mean age 61 years ± 11, 61% males). Most common indications for SPECT were evaluation of ischemic equivalent for coronary artery disease (CAD), risk assessment post-revascularization, and preoperative evaluation for non-cardiac surgery. 10% of referrals were classified as inappropriate, 5% uncertain, and 3% unclassified. Appropriate referrals had a higher proportion of abnormal SPECT results than inappropriate referrals (40% vs 27%, OR 2.08, 95% CI 1.56-2.77, P < .001).

Ward RP, Al-Mallah MH, Grossman GB, Hansen CL, Hendel RC, Kerwin TC, McCallister BD Jr., Mehta R, Dm Polk, Tilkemeier PL,Vashist A, Williams KA, Wolinsky DG, Ficaro EP: American Society of Nuclear Cardiology: American Society of Nuclear Cardiology review of the ACCF/ASNC appropriateness criteria for single-photon emission computed tomography myocardial perfusion imaging

SPECT MPI). J Nucl Cardiol. 2007;14:e26–38.

Gibbons RJ, Miller TD, Hodge D, Urban L, Araoz PA, Pellikka P, McCully RB: Application of appropriateness criteria to stress single photon emission computed tomography sestamibi studies and stress

echocardiograms in an academic medical center. J Am Coll Cardiol. 2008;51:1283–9.

The purpose of this study was to apply published appropriateness criteria for single-photon emission computed tomography (SPECT) myocardial perfusion imaging (MPI) in a single academic medical center.

The study retrospectively examined 284 patients who underwent stress SPECT MPI and 298 patients who underwent stress echocardiography before publication of the criteria.

48% of the inappropriate imaging was in low risk, asymptomatic patients in this study.

**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.*

None

**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**

None

## 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):*

Cardiovascular, Cardiovascular : Coronary Artery Disease

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

Safety : Overuse

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

Elderly, Populations at Risk

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

No current webpage; only NQF specifications page

**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:** Imaging-Efficiency-Measures-Micro-specifications_Measure_Maintenance.doc

**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.

No changes have been made since endorsement.

**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).*

Number of stress SPECT MPI, stress echo, CCTA, and CMR performed for asymptomatic, low CHD risk patients for initial detection and risk assessment*

**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).*

For all orders in asymptomatic patients, determine orders for initial diagnosis and risk assessement. In doing so, patients with known CHD, prior PCI or prior CABG and the following exclusions are not included.

Patients qualify for this numerator if:

- Asymptomatic AND
- Low CHD risk based on clinician estimate AND

NOT any of the following:

- Known CAD, including
- prior MI
- prior ACS
- prior CABG
- prior PCI or
- CHD on prior diagnostic test
- Exercise stress treadmill
- Non-invasive imaging
- Stress echo
- Stress SPECT MPI
- CT Angiography
- Calcium Scoring
- Invasive imaging (cardiac catheterization)
- Ischemic equivalent
- Undergone prior CHD assessment by one the following methods no matter the test result:
  - Exercise stress treadmill
  - Non-invasive imaging
    - Stress echo
    - Stress SPECT MPI
    - CT Angiography
    - Calcium Scoring
  - Invasive imaging (cardiac catheterization)
- Patients for whom preoperative testing is the primary reason for imaging

Submission of individual clinical data variables required for Framingham risk (ATP III criteria) calculation for asymptomatic patients is recognized to place a significant data collection burden upon institutions and may not be possible based on data elements that are readily available at the imaging laboratory. As such, a clinician estimate of CHD risk will be collected for all asymptomatic patients who are being seen for initial detection and risk assessment without known coronary heart disease. However, in making their estimate, clinicians should consider the maximum number of available patient factors used to estimate risk based on Framingham (ATP III criteria), typically age, gender, diabetes, smoking status, and use of blood pressure medication, and integrate age appropriate estimates for missing elements, such as LDL or standard blood pressure. While calculation of the estimate does not require submission of the actual clinical data elements other than the clinician estimate of CHD risk, clinicians are attesting to the accuracy of the estimate by submitting it. An audit of clinician estimates should be completed on a subset of clinicians to verify their estimates as being accurate based on the data that was available.

NOTE: Data collection from patient requisition is required to adequately determine patient's symptom status and clinical risk. Determination with only administrative data is not possible for this measure.

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

Number of stress SPECT MPI, stress echo, CCTA, and CMR performed

**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).*

All consecutive stress SPECT MPI, stress echocardiography, CCTA, and CMR orders

Measurement Entity: Imaging laboratory prospectively measured on test requisition forms and/or patient charts

Level of Measurement/Analysis: Imaging laboratory*

*Attribution for inappropriate use is shared between the ordering physician and imaging laboratory. In an ideal world, attribution to the ordering physician or institution, as well as the imaging laboratory, would be reflected in the reporting of these measures. However, there are numerous complexities that prevent assignment of these measures to individual ordering physicians. For example, ordering volumes from individual physicians and institutions are insufficient to make meaningful comparisons to allow such attribution. Thus, these measures will be reported at the level of the imaging laboratory. However, the extent to which the institution housing the imaging laboratory can impact these measures will be dependent upon cooperation of ordering physicians with the imaging laboratory.

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

None

**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.)*

None.

**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.)*

None.

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

No risk adjustment or risk stratification

If other:

**S.12. Type of score:**

Rate/proportion

If other:

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

Better quality = 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.)*

Locate all stress SPECT MPI, stress echocardiography, CCTA, and CMR orders performed during the sampling period.

Record the total number of tests during the sampling period as the denominator.

From this sets of test orders, identify orders containing the criteria listed in the numerator

**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.

Measures are to be developed based on a sample of a full calendar year based on the following sampling methodology:

Select a starting month:

- o January
- o March
- o May
- o July
- o September
- o November

Begin 60 day data collection period on the 1st on the month for the selected starting month

Determine whether at least 30 stress SPECT, stress echo, CCTA, or CMR orders have been placed during the selected time period. If not, select another time period with a minimum number of 30 cases. If no time period includes the minimum number of cases, then the imaging laboratory does not have sufficient volume to report this measure.

Sampling is required for this measure as full year data collection does not alter performance rates for this measure and would place an additional data collection burden on laboratories. It also allows laboratories to share performance with ordering physicians more quickly than would be possible under full year calendar reporting.

**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.

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

*If other, please describe in S.18.*

Other, Registry Data

**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.

Optimization of Patient Selection for Cardiac Imaging

**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)*

Available in attached appendix at A.1

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

Clinician : Group/Practice, Facility

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

Outpatient Services

If other:

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

**2. Validity – See attached Measure Testing Submission Form**

nqf_testing_attachment_7.1_672_July_2018-636687275323419403_updated.docx

**2.1 For maintenance of endorsement**

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

No

**2.2 For maintenance of endorsement**

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

No

**2.3 For maintenance of endorsement**

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

No - This measure is not risk-adjusted.

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

**Measure Number** (*if previously endorsed*)**:** 672
**Measure Title**: Cardiac stress imaging not meeting appropriate use criteria: Testing in asymptomatic, low risk patients
**Date of Submission**: 11/1/2019

**Type of Measure:**

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

**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: Click here to describe | ☐ other: Click here to describe |

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

N/A

**1.3. What are the dates of the data used in testing**? August 15, 2007 and May 15, 2010

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

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

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

11 practices encompassing 12 ZIP codes within the Chicago metropolitan area; 20 primary care physicians and 2 cardiologists

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.
Circulation. 2013 Oct 8;128(15):1634-43. doi: 10.1161/CIRCULATIONAHA.113.002744. Epub 2013 Sep 10.

**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*)

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Circulation. 2013 Oct 8;128(15):1634-43. doi: 10.1161/CIRCULATIONAHA.113.002744. Epub 2013 Sep 10

Baseline Clinical and Imaging Characteristics

| | Overall Cohort (n=1511) |
|---|---|
| Age, y | 59±13 |
| Women, n (%) | 657 (43.5) |
| Primary indication for MPI, n (%) | |
| Chest pain | 688 (45.5) |
| Dyspnea | 158 (10.5) |
| Abnormal ECG | 136 (9.0) |
| Evaluation of known CAD | 159 (10.5) |
| Preoperative assessment | 37 (2.4) |
| Syncope | 21 (1.4) |
| Asymptomatic | 262 (17.3) |
| Hypertension, n (%) | 841 (55.6) |
| Diabetes mellitus, n (%) | 333 (22.0) |
| Dyslipidemia, n (%) | 695 (46.0) |
| Tobacco use, n (%) | 181 (12.0) |
| Family history of CAD, n (%) | 544 (36.0) |

| | Overall Cohort (n=1511) |
|---|---|
| Framingham 10-y CHD risk, % | 13±10 |
| Likelihood of obstructive CAD, %* | 18±13 |
| Exercise stress (Bruce) protocol, n (%) | 1164 (77.0) |
| BMI, kg/m$^2$ | 30±5.7 |
| Known CAD, n (%) | 271 (17.9) |
| Previous CABG, n (%) | 76 (5.0) |
| Previous PCI, n (%) | 87 (5.8) |
| Previous MI, n (%) | 37 (2.4) |
| Statin, n (%) | 580 (38.4) |
| Antiplatelet, n (%) | 370 (24.5) |
| β-Blocker, n (%) | 307 (20.3) |
| ACE-I or ARB, n (%) | 567 (37.5) |
| Myocardial perfusion, n (%) | |
| Normal (SSS=0–3) | 1344 (88.9) |
| Mildly abnormal (SSS=4–8) | 79 (5.2) |
| Moderately abnormal (SSS=9–13) | 47 (3.1) |
| Severely abnormal (SSS >13) | 41 (2.7) |
| Myocardial ischemia, n (%) | |
| None (SDS ≤1) | 1399 (92.6) |
| Mild (SDS=2–4) | 38 (2.5) |
| Moderate (SDS=5–7) | 40 (2.6) |
| Severe (SDS >7) | 43 (2.8) |
| Type of perfusion abnormality, n (%) | |
| Reversible | 87 (5.8) |
| Fixed | 61(4.0) |
| Reversible and fixed | 19 (1.3) |

| | Overall Cohort (n=1511) |
|---|---|
| Poststress LVEF <50%, n (%) | 78 (5.2) |

**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 cohort used for the validity testing is described above. A smaller single center study was used for reliability testing and is cited below.

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging.

The demographics of the single center study are as follows: May 1, 2005, and May 15, 2005. Mayo Clinic (Rochester, Minn). The mean±SD age of the 298 study patients was 66±13 years; 52% were men, 20% had diabetes mellitus, 60% had hypertension, 66% had hyperlipidemia, 54% had a history of smoking, 11% had a prior myocardial infarction, 20% had prior coronary revascularization, 36% had chest pain, 38% had dyspnea, and 41% had a normal resting ECG.

**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.

N/A

---

**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*)

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging.

Using the appropriateness criteria document, 2 experienced cardiac registered nurse abstractors reviewed patient demographics and other relevant information and classified each patient as appropriate, inappropriate, or uncertain. Patients who did not fit any of the clinical situations in the appropriateness criteria were judged to be not classifiable. The level of agreement between the 2 raters was analyzed. Patients who did not fit the measure were deemed unclassified as they did not conform to the available scenarios. It does not imply that data was unavailable to determine the appropriateness of scenarios that had been published, including the focus of this measure.

Also, see section 2b1 for validity testing of data elements.

**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*)

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging. 2009 May;2(3):213-8. Nurse abstracter agreement kappa=0.72 for stress echocardiography

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

The data elements required for calculation of the appropriate use metrics can be obtained reliably by clinical staff from data residing in patient records with a high degree of agreement between nurses who would enter the data into the registry/clinical database.

---

**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)*

Relationship between appropriate use score and predictive value of SPECT MPI

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

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Patients in the appropriate/uncertain group experienced significantly higher overall rates of death (HR, 2.9; 95% CI, 1.05–8.0; P=0.04), the composite of death or MI (HR, 1.04; 95% CI, 1.01–1.07; P=0.03), and the composite of cardiac death or MI (HR=5.7; 95% CI, 1.3–25.6; P=0.02) after adjustment for clinical covariates. Among patients in the appropriate/uncertain group, abnormal MPI continued to predict a multifold increase in the risk of death, cardiac death, composite of death or MI, and composite of cardiac death or MI (Figure 4). However, in the inappropriate group, there were no statistically significant differences in MACE rates between subjects with abnormal versus normal MPI (Figure 4). Furthermore, using Cox regression models, no interaction was identified between the study group and MPI finding in predicting death, the composite of death or MI, or the composite of cardiac death or MI (P=0.91, 0.70, and 0.43, respectively).

A Cox regression model demonstrated that inappropriate MPI use was a negative predictor of all-cause mortality (HR, 0.26; 95% CI, 0.10–0.67; P=0.005) after adjustment for myocardial perfusion finding (normal versus abnormal; HR, 2.5; 95% CI, 1.1–5.9; P=0.04) and depressed LVEF (<50%; HR, 3.7; 95% CI, 1.5–9.3; P=0.006); undergoing early coronary revascularization was not predictive of mortality (P=0.98). Similarly, in separate models, we demonstrated that inappropriate use was an independent negative predictor of the secondary end points of death or MI (HR, 0.31; 95% CI, 0.14–0.70; P=0.005) and cardiac death or MI (HR, 0.16; 95% CI, 0.04–0.71; P=0.02) after adjustment for depressed LVEF, myocardial perfusion findings, and early revascularization. In these models, MPI and depressed LVEF independently predicted the composite end points of death or MI and cardiac death or MI, whereas undergoing early coronary revascularization after MPI was not predictive of these end points (P≥0.97). Finally, in forward stepwise Cox regression models, appropriate use was shown to have incremental prognostic value to perfusion imaging and depressed LVEF in predicting MACE; undergoing early revascularization (<60 days) did not provide significant additional predictive value

**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?*)

Appropriateness of imaging as measured by these metrics is correlated with the downstream value of the test in contributing to clinical decision making. As such, the metrics contribute to ensuring the prognostic value of the imaging procedures measured.

_____

**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*)

**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*)

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

---

**2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES**
***If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.***

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **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.**

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

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

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

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

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

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

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

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

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

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

**2b3.9. Results of Risk Stratification Analysis**:

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

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

---

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

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

To date, there has been consistency in the studies showing the gap in performance on these metrics across sites. While individual practitioner level measurement and rates based on type of physician have shown variability, practice/hospital performance has been similar at baseline and after intervention to improve.

**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*)

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

No statistical tests have been applied to demonstrate differences among the measured entities at the practice/hospital level. Inappropriate use rates for individual practitioners ranged from 10% to 77% (P<0.001) and were higher among primary care physicians than cardiologists (47% versus 28%; P<0.001)

Fonseca R, Negishi K, Otahal P, et al. Temporal Changes in Appropriateness of Cardiac Imaging. J Am Coll Cardiol. 2015 Mar 3;65(8):763-73.

A separate meta-analysis demonstrated wide variation of appropriate use rates as described in the performance scores over time in section 2.b.1.

**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?*)

Inappropriate use is common among a wide range of practices and hospitals. Variability exists within practices and provides opportunities for peer to peer learning and improvement on these measures, especially within a practice or between primary care and specialists.

---

**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*)

**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*)

**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*)

---

**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*)

All subjects were classifiable according to the 2009 AUC and therefore no analysis for missing data was required. For validity testing, some subjects were lost to follow-up. Their demographic and AUC patterns were analyzed for similarity to the included patient cohort.

**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*)

Doukky R, Hayes K, Frogge N, Balakrishnan G, et al. Impact of appropriate use on the prognostic value of single-photon emission computed tomography myocardial perfusion imaging.

Compared with subjects with complete follow-up, the patients excluded (n=182) or lost to follow-up (n=14) were younger (mean age, 55±15 versus 59±13 years; P=0.001) and had lower likelihood of obstructive CAD (15±13% versus 18±13%; P=0.007) but similar mean 10-year Framingham coronary heart disease risk (12.7±10.8% versus 12.8±10%; P=0.88) and CAD prevalence (19% versus 18%; P=0.62). The prevalence of depressed LVEF and abnormal perfusion was nearly identical (P=0.97 and 0.89, respectively), with a similar breakdown of reversible, fixed, and mixed defects (P=0.64). The excluded patients had a similar distribution of AUC classifications: 104 (53.1%) appropriate, 89 (45.4%) inappropriate, and 3 (1.5%) uncertain (P=0.53).

**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*)

Performance results were not biased as no missing data was recorded for the metrics themselves. Validity testing showed similar distribution of AUC, perfusion defects, and CAD prevalence and thus unlikely to have impacted results of this testing.

## 3. Feasibility

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

**3a. Byproduct of Care Processes**

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

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

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Other

If other: An EHR or Web portal prompts for clinical information in a decision support tool for individual cases that then are transmitted to a measurement registry

**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**.

Some data elements are in defined fields in electronic sources

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

Some data elements should already be a part of the electronic record (PCI history, scheduled surgery). In addition, e-ordering for diagnostic testing has been proposed for meaningful use, encouraging integration of these types of data elements. In addition, ACC has developed clinical decision support tools that can be embedded in electronic health records to capture the necessary information.

**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.**

Hendel, RC; Cerqueira, M; Douglas, PS et al. "A Multicenter Assessment of the Use of Single-Photon Emission Computed Tomography Myocardial Perfusion Imaging With Appropriateness Criteria". J Am Coll Cardiol. Published online December 10, 2009.

This study demonstrated the feasibility of data collection as well as the most frequent inappropriate indications. This allowed ACC to narrow the number of indications measured for this measure set along with the associated data elements.

**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)*.

None required. Decision support tools are available to aid in data collection and are available on a per test basis.

## 4. Usability and Use

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

**4a. Accountability and Transparency**

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

**4.1. Current and Planned Use**

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

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|
| | Public Reporting<br>PQRS<br>http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html<br>Payment Program<br>QPP<br>https://qpp.cms.gov/mips<br>QPP<br>https://qpp.cms.gov/mips<br>Regulatory and Accreditation Programs<br>IAC<br>http://www.intersocietal.org/intersocietal.htm<br>Professional Certification or Recognition Program<br>FOCUS<br>www.cardiosource.org/focus<br>Quality Improvement (Internal to the specific organization)<br>FOCUS<br>https://www.acc.org/focus |

**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

QPP/MIPS - CMS/pay for performance/national; The data collected at the lab level for this measure can be further segmented by physician to help them understand their appropriate use patterns; although small sample sizes can limit comparability for some providers.
FOCUS - ACC/lab accreditation, quality improvement and utilization management/national - 25,000 cases with concentrations in DE (100% for SPECT MPI) and Western PA (10% for SPECT MPI and stress echo for cardiologists) - addtional 6,000 cases
IAC - lab accreditation/national - 100% - 5% of lab tests performed on an annual basis

**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.**

through CMS

**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.**

through CMS

**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.**

**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.**

**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.**

None have been identified at this time.

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

## 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.

No

**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.**

**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?**

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

**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.)**

## Appendix

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

Attachment **Attachment:** FOCUS_Data_Collection_Sheet-635249624195073013.docx

## Contact Information

**Co.1 Measure Steward (Intellectual Property Owner):** American College of Cardiology

**Co.2 Point of Contact:** Amy, Dearborn, adearborn@acc.org, 202-375-6257-

**Co.3 Measure Developer if different from Measure Steward:** American College of Cardiology Foundation

**Co.4 Point of Contact:** Joseph, Allen, jallen@acc.org, 202-375-6463-

## 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.**

All individuals are volunteer members representing American College of Cardiology Foundation:

Pamela Douglas, MD, MACC

Joseph Allen, MA

Robert Hendel, MD, FACC

Joseph Cacchione, MD, FACC

Manuel Cerqueira, MD, FACC

Joseph Drozda, MD, FACC

Michael Picard, MD, FACC

Martha Radford, MD, FACC

Leslee Shaw, PhD, FACC

Allen Taylor, MD, FACC

Group developed list of proposed measures, specifications, definitions, justification, etc.

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 2009

**Ad.3 Month and Year of most recent revision:** 11, 2019

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

**Ad.5 When is the next scheduled review/update for this measure?** 12, 2020

**Ad.6 Copyright statement:** Copyright 2009. American College of Cardiology Foundation

**Ad.7 Disclaimers:**

**Ad.8 Additional Information/Comments:** date above refers to maintenance of endorsement

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 #:** 3534

**Corresponding Measures:**

**De.2. Measure Title:** 30 Day All-cause Risk Standardized Mortality Odds Ratio following Transcatheter Aortic Valve Replacement (TAVR).

**Co.1.1. Measure Steward:** American College of Cardiology

**De.3. Brief Description of Measure:** This measure estimates hospital risk standardized odds ratio for death from all causes within 30 days following transcatheter aortic valve replacement. The measure uses clinical data available in the STS/ACC TVT Registry for risk adjustment. For the purpose of development and testing, the measure used site-reported 30-day follow-up data contained in the STS/ACC TVT Registry.

**1b.1. Developer Rationale:** This measure will describe hospital-level 30-day mortality rates following TAVR, with the overriding goal to reduce 30-day mortality rates. The expectation is that providing this information to hospitals, coupled with public reporting of hospitals' results, will drive internal hospital quality improvement efforts to focus efforts on reducing TAVR mortality. Of note, the measure includes in-hospital deaths and deaths occurring after hospital discharge up to 30 days post procedure. This perspective may motivate hospitals to look for opportunities not only within the organization, but to better coordinate the transition of care from the inpatient to the outpatient arena.

**S.4. Numerator Statement:** The outcome of this measure is all-cause death within 30 days following a transcatheter aortic valve replacement (TAVR).

**S.6. Denominator Statement:** The target population for the outcome is for individuals who have undergone transcatheter aortic valve replacement.

For development, reassessment and reporting of this measure, we use site reported data from the STS/ACC TVT Registry.

**S.8. Denominator Exclusions:** 1)Hospitals need to meet eligibility criteria to be included in the measure.

2) Patients are excluded if:

a) They did not have a first-time TAVR in the episode of care (admission),

b) The TAVR was subsequent to another procedure in the Registry (other TAVR, Mitral Leaflet Clip and/or TMVR) during that admission.

c) The patient is readmitted for a repeat TAVR (re-admission) and the initial TAVR was performed during the rolling 3-year timeframe for the measure.

d) 30-day mortality status missing.

**De.1. Measure Type:** Outcome

**S.17. Data Source:** Registry Data

**S.20. Level of Analysis:** Facility

## Preliminary Analysis: New Measure

### Criteria 1: Importance to Measure and Report

#### 1a. Evidence

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

**Evidence Summary**

The developer presents evidence for two factors they state are within a hospital's control and can improve 30-day mortality rates: appropriate patient selection and volume of TAVR.

- One study analyzed data from the TVT Registry and found patients with very poor health status had a two-fold increased hazard of death over the first year after TAVR as compared to patients with good health status. Patients with poor and fair health status had intermediate outcomes. The analysis adjusted for a broad range of baseline covariates.
- Two studies examined the relationship between volume and mortality, vascular complications, and stroke. The authors noted mortality at 30 days was higher and more variable at hospitals with a low procedural volume than at hospitals with high procedural volume.

***Question for the Committee:***

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

**Guidance from the Evidence Algorithm**

Outcome measure (Box 2) → Link between outcome and healthcare action → Pass

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

#### 1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

**1b. Performance Gap.** The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer shared analysis from the TVT Registry for two partially overlapping 3-year time periods. The Risk Standardized Odds Ratio is calculated as the odds that an outcome (e.g. 30-day mortality) will occur for patients treated at a facility compared to the "odds" that outcome will occur for patients with identical risk factors if treated

by a hypothetical (average) hospital. Thus, a lower odds ratio implies lower-than-expected mortality (better quality) and a higher ratio implies higher-than-expected mortality (worse quality).

The results show variation in performance. The variation increases from the older to the more recent time period.

TVT Registry data June 2013 – May 2016 (21,661 TAVR patients from 188 TVT hospitals)

Distribution of hospital-specific odds ratio estimates

| **Mean** | **Std Dev** | **Min** | **10th** | **20th** | **25th** | **30th** | **40th** | **50th** | **60th** | **70th** | **75th** | **80th** | **90th** | **Max** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.00 | 0.02 | 0.92 | 0.97 | 0.98 | 0.99 | 0.99 | 1.00 | 1.00 | 1.01 | 1.01 | 1.01 | 1.02 | 1.03 | 1.07 |

TVT Registry data April 2015 – March 2018 (49,182 TAVR patients from 265 TVT hospitals)

Distribution of hospital-specific odds ratio estimates

| **Mean** | **Std Dev** | **Min** | **10th** | **20th** | **25th** | **30th** | **40th** | **50th** | **60th** | **70th** | **75th** | **80th** | **90th** | **Max** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.01 | 0.10 | 0.81 | 0.89 | 0.94 | 0.95 | 0.96 | 0.98 | 1.00 | 1.02 | 1.04 | 1.06 | 1.07 | 1.13 | 1.40 |

**Disparities**

The developer states, "In order to explore disparities, we modified the measure´s hierarchical model to include indicator variables for black race, other non-white race, Hispanic ethnicity, and participation in Medicaid. We performed this analysis using data from June 2013 to May 2016 (21,661 patients from 188 hospitals) and using data from April 2015 to March 2018 (49,182 patients from 264 hospitals). In order to accommodate these variables, we removed an existing related variable that was defined as "non-white race or Hispanic ethnicity". Results are summarized in the form of odds ratios below. For each variable in each time period, the 95% confidence interval around the odds ratio overlaps with the null value of 1.0. This implies that there was no statistically significant association between these variables and 30-day mortality after adjusting for other factors in the hierarchical model ($p>0.05$ for each variable below)."

| **Variable** | **June 2013 – May 2016** | **April 2015 – March 2018** |
|---|---|---|
| Medicaid | 0.93 (0.69 – 1.24) | 1.05 (0.83 – 1.32) |
| Black race (versus white) | 0.73 (0.47 – 1.14) | 0.91 (0.67 – 1.23) |
| Other non-white race (versus white) | 0.68 (0.36 – 1.30) | 0.85 (0.54 – 1.34) |
| Hispanic ethnicity | 1.23 (0.82 – 1.84) | 0.82 (0.59 – 1.16) |

***Questions for the Committee:***

- Is there a gap in care that warrants a national performance measure?

**Preliminary rating for opportunity for improvement: ☐ High ☒ Moderate ☐ Low ☐ Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)**

1a.

- I agree with the preliminary rating of pass
- Mortality is an ultimate outcome measure and in this setting due to operator/institution skill and experience AND patient selection.
- Strong data to support this measure
- 30 day all cause mortality - assumes that better patient selection or increases in volume will improve. No systematic review. 3 articles cited. At least one other study found related to volume and decrease mortality. Evidence pass but low

1b.

- The performance gap is wider in the 2015-18 cohort than the 2013-16 cohort. This may be due to the fact that mortality rates are falling and some hospitals are being left behind. Also,more hospitals doing the procedure or an unidentified covriate. Disparities are analyzed but no gaps are found. CIs are surprisingly wide Disparities are analyzed but no gaps are found. CIs are surprisingly wide
- adequate evidence of a gap
- Looking at the old data, not sure there is a gap; however, the StnDev is larger in the newer data, which has a bigger gap.
- There was a performance gap in the most recent 3-year period. Disparities testing yielded no significant disparities.

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability: Specifications and Testing**

**2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data**

**2c. For composite measures: empirical analysis support composite approach**

### Reliability

**2a1. Specifications** requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

**2a2. Reliability testing** demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

### Validity

**2b2. Validity testing** should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

**2b2-2b6. Potential threats to validity** should be assessed/addressed.

**Composite measures only:**

**2d. Empirical analysis to support composite construction**. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

**Complex measure evaluated by Scientific Methods Panel**? ☒ **Yes** ☐ **No**

**Evaluators:** NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

**Methods Panel Evaluation Summary**:

This measure was reviewed by the Scientific Methods Panel and discussed during the October 2019 In-Person Meeting. The Subgroup did not pass the measure on reliability. The Subgroup was unable to reach consensus regarding validity. During the in-person meeting, the full Panel discussed reliability and validity and then the Subgroup voted again, passing the measure on both reliability and validity. A summary of the measure and the Panel discussion is provided below.

Reliability (final vote) 0-H; 6-M; 0-L; 0-I → Moderate reliability

- To demonstrate reliability of the data elements used in the measure, the developer assessed inter-rater reliability using data from 40 records selected randomly from 4 randomly selected facilities (presumably, 10 records per facility, although this is not clear). The Subgroup initially rated the measure low for reliability.
- Key concerns in the initial analysis related to reliability were the lack of detail around the testing and sampling methodology and that not all data elements were evaluated for reliability (or validity).
- In response to the concerns raised, the developer provided additional information regarding the sampling, demonstrating no systematic patient differences between those selected for sampling and the general cohort and provided IRR results for additional data elements. On re-vote, the measure passed validity with a moderate rating.

Validity (final vote) 0-H; 5-M; 1-L; 0-I → Moderate validity

- To demonstrate validity of the data elements, the developers conducted 2 analyses:
    - Record eligibility assessment: 6 hospitals participating in the registry reported all TAVT and Mitral cases performed at their facility during a specified timeframe. These were compared to those records included in the registry to verify that cases were not missed. N=366 records
    - 40 hospitals with at least 10 cases were randomly selected for audit. From each hospital, 10 baseline and 10 follow-up cases (for 30-day and 1-year) were randomly selected for abstraction. Sample included 400 “baseline” records, 400 “30-day” records, and 289 “1-year” records. Developers calculated the prevalence-adjusted and bias-adjusted kappa (PABAK) statistic.
- Key concerns in the SMP’s initial analysis regarding the measure include exclusion of >50% of hospital/patients due to missing data, relatively low values of PABAK for two tested values, lack of data element testing for most variables, and a relatively small testing sample that may or may not be representative of hospitals/patients included in the measure. The Subgroup was initially unable to reach consensus on validity.

- In response to the concerns raised, the developer provided additional information regarding key data elements and thresholds for excluding hospitals/patients. The developer also performed validity testing on additional data elements. The developer defended keeping baseline KCCQ-12 and baseline gait speed in the data model, indicating they anticipate more sites will complete these elements because they are required for the measure. They feel both elements are clinically important for patient evaluation. On re-vote, the measure passed validity with a moderate rating.

***Questions for the Committee regarding reliability:***

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

***Questions for the Committee regarding validity:***

- During the SMP discussions, questions were raised about the trade-off between keeping baseline KCCQ-12 and baseline gait speed in the risk adjustment. The developer feels these are clinically important items to include; however, currently not all hospitals have this information available, resulting in hospitals being excluded for missing data.
- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

**Preliminary rating for reliability:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

**Preliminary rating for validity:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)**

2a1.

- The measure was rated and passed as moderate by the sci methods panel
- Don't like exclusion of patients missing 30 day mortality indicator
- Was reviewed by Scientific Methods. Aligned with their work
- Detailed specifications of registry items for risk adjustment. Death in registry compared to CMS data with 99.6% agreement

2a2.

- >50% of hospitals or patients were excluded due to missing data. Apparently the missing data are usually KCCQ-12 and gait speed.
- Adequate
- Was reviewed by Scientific Methods. Aligned with their work
- Inter-rater reliability was good overall 97.7%

2b1.

- The SMP had difficulty with accepting validity.
- no and agree with developer that KCCQ and gait test are important to leave in
- Was reviewed by Scientific Methods. Aligned with their work

- Some problems noted with one particular site around missing data for the outcome indicator of death (40 hospitals randomly chosen for validity testing) Empirical validity of data elements ranged from 0.63 to 1.00 indicating substantial agreement

2b4-7.

- Gait speed data and KCCQ-12 scores may be missing at high rates despite being required for the measure
- Missing 30 day mortality field does threaten validity
- Was reviewed by Scientific Methods. Aligned with their work
- Missing data seemed to be an issue in only one site

2b2-3

- Again, gait speed data and KCCQ-12 scores may be missing frequently
- As TAVR is now being performed in multiple risk groups risk adjustment is needed, I do not know if the registry includes enough information to allow this, but suspect it does or shortly will.
- Was reviewed by Scientific Methods. Aligned with their work
- There was support for the risk adjustment via cited article for the development but the contributions of each variable to the risk adjustment were not provided.

## Criterion 3. Feasibility

**Maintenance measures – no change in emphasis – implementation issues may be more prominent**

**3. Feasibility** is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer states that all data elements associated with this measure are routinely generated and acquired during the delivery of standard cardiac care to this patient population, with the exception of the Kansas City Cardiomyopathy Questionnaire and six-minute walk test.
- The developer states a full-time employee can enter roughly 1,200 patient records per year on average.
- The developer states that all hospitals performing TAVR participate in the registry as a condition of a CMS coverage with evidence decision.

***Questions for the Committee:***

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?

**Preliminary rating for feasibility:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

- Kansas City Cardiomyopathy Questionnaire and six-minute walk test are not routinely collected. All hospitals doing TAVR must participate in the registry as per CMS.
- CMS mandated use of registry ensures feasibility
- More details on KCCQ-12 and gait would be beneficial based on SMP and NQF staff questions

- Yes - registry that is required by CDC

## 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**

The measure is used as part of ACC's Transcatheter Valve Certification program and measure results are used to quality improvement purposes by registry participants. In the future, STS and ACC plan to publicly report the measure results.

**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

- The developer states that each participant receives quarterly feedback reports providing a detailed analysis of the participant's performance including benchmarking. Participants also have access to a guide to help interpret performance results.

**Feedback on the measure by those being measured or others**

- The developer reports that feedback is typically obtained through monthly registry site manager calls, ad hoc calls, and break-out sessions at the registry's annual meeting. They report feedback has generally been supportive and positive regarding the measure and registry. No changes have been made to the measure at this time.

**Additional Feedback:**

***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**

## 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**

Between 2014 and 2017, the aggregate 30-day TAVR mortality rate in the analysis population decreased from 5.9% to 2.7%, representing a relative decrease of 54%.

Overall 30-day Mortality:

2014: 5.9%

2015: 4.2%

2016: 3.1%

2017:2.7%

The developer estimates that some improvement is due to a shift in case mix to lower-risk patients but that some improvement is due to quality improvement efforts by facilities. "In the hierarchical logistic regression model for the time period June 2013 to May 2016 accounting for differences in case mix, the estimated odds of mortality decreased 15% per year (odds ratio per year 0.85, 95% CI 0.78 to 0.93, p<0.001), which is a more appropriate estimate of improvements in care at a facility level."

**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 reports no unexpected findings or unintended consequences.

**Potential harms**

None noted.

**Additional Feedback:**

***Questions for the Committee**:*

- Are you aware of any unintended consequences related to this measure?
- Do the benefits of the measure outweigh any potential unintended consequences?

**Preliminary rating for Usability and use:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 4: Usability and Use**

4a1.

- The measure is used as part of ACC's Transcatheter Valve Certification program and measure results are used to quality improvement purposes by registry participants. In the future, STS and ACC plan to publicly report the measure results.

- Not clear that this measure adds anything to the current registry reporting
- Concerned measure is not used publicly. The evidence/rationale is based on the measure results causing peer recognition and improvement.
- Feedback was solicited and generally positive. No changes were made.

4b1.

- No significant harms identified
- As with any public reporting of mortality, there is a risk that programs will deny care to the sickest patients to keep their mortality rate down. The measure's use of an O/E ratio does obviate this to a large degree.
- No concerns
- Benefits outweigh the risks

## Criterion 5: Related and Competing Measures

**Related or competing measures**
2561: STS Aortic Valve Replacement (AVR) Composite Score
**Harmonization**
The measures have been harmonized to the extent possible.

**Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures**

- 2561: STS Aortic Valve Replacement (AVR) Composite Score. The two measures are harmonized to the extent possible.
- Not sure how the measure adds to the existing registry that is reported out.
- Harmonized as much as possible
- 2561 Aortic Valve Replacement Composite score - developers say different populations

## Public and Member Comments

**Comments and Member Support/Non-Support Submitted as of: Month/Day/Year**

- **Of the XXX NQF members who have submitted a support/non-support choice:**
  - XX support the measure
  - YY do not support the measure

Combined Methods Panel Scientific Acceptability Evaluation

**Measure Number:** 3534

**Measure Title:** 30 Day All-cause Risk Standardized Mortality Odds Ratio following Transcatheter Aortic Valve Replacement (TAVR).

**Type of measure:**

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☒ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

**Data Source:**

☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☒ **Registry Data**
☐ **Enrollment Data** ☐ **Other**

**Level of Analysis:**

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

**Measure is:**

☒ **New** ☐ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

## RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   **Submission document:** "MIF_xxxx" document, items S.1-S.22

   ***NOTE****: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

   **Panel Member #1:** No concerns.

   **Panel Member #2:** The specifications are very context specific with respect to participation in the STS/ACC TVT Registry. It is doubtful that any organization outside the registry or independent of the registry would be able to replicate the specification. So in that sense reliability of the specification is somewhat difficult to ascertain. Also the 30-day mortality "proxy" should be independently validated (by applying the same 75-day specification for the cases with no missing data).

   **Panel Member #5:** The measure is based on elements contained in an existing registry, only sites with high levels of completeness will have scores calculated, and the measure will be calculated by the measure sponsor. Therefore, consistency of implementation does not seem to be a cause for concern.

   **Panel Member #6:** *No concerns.*

   **Panel Member #7:** I have no concerns. The specifications and concise and unambiguous.

## RELIABILITY: TESTING

**Submission document:** "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☐ **Measure score** ☒ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☒ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☒ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**

   **Submission document:** Testing attachment, section 2a2.2

   **Panel Member #1:** The method used is appropriate but not comprehensive. Also, it may be more accurate to name it as a percent agreement test rather than an inter-rater reliability test.

   Only 4 out of over 40 data elements needed to calculate the measure score were tested for agreement. This does not represent the full data extraction process, thus there is no testing of the agreement/reliability of the majority of data elements gathered from the registry. If they were tested previously when creating the registry, information on these tests and results should be provided or referenced.

   Also, no comparison was made between the characteristics of patients from the 40 records tested and patients from the full dataset to test their representativeness. It is possible that records of patient with different levels of medical complexities may differ in the difficulty of data extraction.

   **Panel Member #2:** One concern is the reliance upon 40 records (out of 21,000) to estimate data element reliability, especially since apparently none of the randomly selected cases had the outcome of interest (mortality). Also the focus seems to be on the facility level eligibility criteria (e.g. baseline 5 meter walk test performed) rather than measure data elements.

   **Panel Member #3:** Two auditors performed inter-rater abstracted 40 charts, and inter-rater reliability was assessed, which showed perfect agreement.

   **Panel Member #4:** *IRR for 40 patients across 4 facilities on 6 variables (one of which was follow up date of death and was noted as "N/A" for all).*

   **Panel Member #5:** The description of the reliability testing was a little unclear. "Two trained auditors performed inter-rater reliability (IRR), performing a visual inspection of the medical record for the sample cases (each reviewed the records in the sample) to abstract necessary data. IRR assessment was performed on baseline, and 30-day and one-year follow-up cases." It appears that the IIR was calculated by comparing the two auditors' independent abstracts of the same records. It is important to know how the abstractors were picked and trained. How do we know that they are representative of abstractors at the other >400 sites? Also, it is concerning that only 6 data elements were tested and that the 40 cases came from only 4 sites. What about all of the other risk model inputs and sites? Also, none of the selected cases experienced 30-day mortality. The description of the method implies that multiple time points were assessed but it's not clear if reliability statistics for more than one timepoint are reported.

   **Panel Member #6:** *40 records from 4 facilities were chosen at random and two trained auditors (presumably independently) to abstracted necessary data from the medical record. IRR was calculated at baseline, and 30-day and one-year follow-up for six data elements.*

   **Panel Member #7:** The steward assessed interrater reliability of data elements. Two raters were compared.

7. **Assess the results of reliability testing**

   **Submission document:** Testing attachment, section 2a2.3

**Panel Member #1:** The 100% agreement achieved might not be representative of true agreement due to the limitations noted above.

**Panel Member #2:** For those facility level eligibility criteria registry data elements that were assessed the agreement was 100% (although is "agreement rate" the right statistic?)

**Panel Member #3:** Sample size too small and limited to too small a subset of the data elements to adequately assess reliability of data elements (only 40 records were evaluated). However, data element validity, assessed by re-abstracting the data elements and comparing them to the registry data elements was performed in a much larger data set.

| Data Element | Numerator | Denominator | Agreement Rate |
|---|---|---|---|
| Discharge Status (DCStatus) | 40 | 40 | 100.0% |
| Discharge Date (DCDate) | 40 | 40 | 100.0% |
| Follow-up Status (F_Status) | 39 | 39 | 100.0% |
| Follow-up Date of Death (F_DeathDate) | 0 | 0 | N/A |
| Five Meter Walk Test Performed (FiveMWalkTest) | 36 | 36 | 100.0% |
| KCCQ-12 Performed (KCCQ12_Performed) | 40 | 40 | 100.0% |

**Panel Member #4:** *This is sparse testing, however, the STS and its partner on this measure describe comprehensive data auditing in their registry that may be considered a proxy for all data reliability.*

**Panel Member #5:** Although the calculated IIR were 100% for 5 of the 6 variables tested, many of the risk model inputs were not tested. The sample size was very small for cases and sites with no variability for the key outcome variable. As noted, the methodological concerns stated above make the results hard to interpret.

**Panel Member #6:** *Although there was complete agreement among the auditors, it is hard to know whether this would extend to other hospitals or data elements.*

**Panel Member #7:** Agreement between the two raters was complete (100%).

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**

   **Panel Member #1:** Not all data elements were tested, only a small number of elements were assessed.

   ☐ **Not applicable** (data element testing was not performed)

   **Panel Member #4:** *I am interested in others' impressions of this. As a stand alone, the offered assessment of data reliability (40 patients with a few data points, all of those being reported as 100% IRR or NA) seems sparse.*

*This may be counterbalanced by the extensive, established data integrity for which STS is known, but I seek others' perspectives. The subsequent *validity* estimates (PABAK scores from 40 hospitals and about 400 patients, 6 variables) are more compelling.*

**Panel Member #6:** *Only six elements were assessed.*

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☒ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☒ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    **Panel Member #1:** The limitations noted above could create a false impression of high accuracy of data element extraction. A more comprehensive testing protocol including all data elements would yield more comprehensive results, and provide the information needed to rate the measure's reliability.

    **Panel Member #2:** Given that the data element reliability demonstration did not include the outcome of interest the demonstration is borderline insufficient, and the methods are borderline inadequate.

    **Panel Member #3:** Reliability testing was performed on a small sample of charts, and can defended on the basis of the cost of data abstraction. However, this limited data reliability testing is not sufficient to establish the overall reliability of the measure. But, the high level of data validity suggests that the data is reliable. However, the Measure developers should evaluate the reliability of the measure score, since this is not resource-intensive and relatively straightforward to do.

    **Panel Member #4:** *Provisionary assessment: Explanation as above, this assessment of moderate may be generous given the information provided.*

    *One site had 0% agreement on follow-up status. This may be concerning given an expectedly low event rate.*

    **Panel Member #5:** Item-level reliability testing was not presented for many of the risk model elements. The low sample size (40), zero event rate for death, and use of only 2 raters of unknown providence make the analyses of reliability hard to interpret.

    **Panel Member #6:** *I am concerned that the small sample of hospitals and data elements is not sufficient to assess reliability.*

    **Panel Member #7:** Data element reliability is high, although with assessment of only 40 cases. However, score reliability is unknown.

## VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    **Submission document:** Testing attachment, section 2b2.

    **Panel Member #1:** No concerns. However, a table demonstrating the exact number of sites and patients excluded due to each exclusion criteria would be informative.

**Panel Member #2:** Testing results are always context specific, and in this case the context includes participation in the registry and meeting the facility eligibility criteria. The testing results (reliability and validity) are not applicable and should not be used as the basis of a decision outside that context.

**Panel Member #3:** none

**Panel Member #4:** *None.*

**Panel Member #5:** The measure will not be calculated for facilities that have <90% complete data on selected variables. Less than half (188 of 450) hospitals met this criterion (253 of 301 in later years. Although this exclusion probably increased the validity of the measure for the included hospitals, most hospitals will not be able to participate in the measure. The rationale and description of the limitations of this approach are well-stated in the application. The rationale for selection of the exclusion variables (30-day status, baseline KCCQ-12 score, and baseline gait speed) was not given. What about missing data on other key variables?

**Panel Member #6:** *It is hard to know what effect exclusions have on the measure's validity. Only hospitals with relatively complete data are included in the assessment, and the results presented in 2b2 describe differences in the hospitals included and not rather than differences in the measures in these hospitals. Consequently, the results presented say nothing about the effect of exclusions on the final measures, especially if calculated for hospitals with high levels of missingness.*

**Panel Member #7:** The exclusion of hospitals with less than 90% complete, non-missing data is impactful. Excluded hospitals are more likely to be teaching hospitals, to be larger, and to have more minority patients. Because the measure is a standardized odds ratio, exclusion of a large number of hospitals alters the frame of inference, thus raising concern about external validity. (That is, how does mortality associated with TAVR at an included hospital compare to mortality at an excluded hospital?)

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

**Submission document:** Testing attachment, section 2b4.

**Panel Member #1:** No concerns.

**Panel Member #2:** A better analysis might report the percentage of facilities with a predicted odds less than 1.0 and greater than 1.0 as some specified degree of confidence.
**Panel Member #3:** none
**Panel Member #4:** *N/A.*
**Panel Member #5:** The distribution of performance presented (histogram) is convincing. A catapiller plot with CIs, would be even more informative.

**Panel Member #6:** *No concerns.*

**Panel Member #7:** The distribution of standardized odds ratios is relatively narrow. Most values are between 0.9 and 1.1. The steward presents no information about the confidence intervals associated with each standardized odds ratio. Thus, it is unclear whether any of the estimated odds ratio indicate that mortality is significantly better or worse than is expected, given risk adjustment.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

**Submission document:** Testing attachment, section 2b5.
**Panel Member #1:** NA
**Panel Member #2:** Not applicable.
**Panel Member #4:** *N/A.*

**Panel Member #5:** The decisions, rationale, and methods regarding missing data are reasonable, strategic, and pragmatic.

**Panel Member #6:** *NA.*

**Panel Member #7:** This item is not applicable.

15. **Please describe any concerns you have regarding missing data.**

**Submission document:** Testing attachment, section 2b6.

**Panel Member #1:** The majority of sites and patients were excluded due to the exclusion criteria of <90% completeness on 30 day mortality status, KCCQ-12, or gait speed, with the aim of reducing bias. Obviously, this large exclusion rate introduces a potential for a selection bias of itself.

As described in the submission, the imputation of missing KCCQ-12 and gait speed to the median can in this case slightly correct for bias due to missing data as imputation would make patients with missing data look healthier, therefor would 'benefit' less from risk-adjustment. Additional more robust methods for adjusting for patient censoring as inverse-probability-weighting (IPW) might be a better option. IPW would correct for all available variables associated with the probability of having missing data, not just those that have missing data that were imputed, which could in some cases increase the selection bias, depending on the difference between patients with complete or incomplete data.

Another option would be to reconsider the high inclusion threshold of ≥90% of complete data which resulted in an average of 96% complete data on 30 day mortality. Maybe a threshold that would set included sites at, for example, an average of 80- 90% completeness (instead of 96%) would be more reasonable, enabling more sites to be included while still maintaining a high completion rate.

**Panel Member #2:** Testing results are always context specific, and in this case the context includes participation in the registry and meeting the facility eligibility criteria. The testing results (with facilities and patients with missing data excluded) are not applicable and should not be used as the basis of a decision outside that context.

**Panel Member #3:** The initial cohort consisted of 60,770 records from 450 hospitals. The final cohort after excluding patients hospitals with >90% of missing data was 21,161 records from 188 hospitals. The "key" data elements included KCCQ-12 score and baseline gait speed. Although the final model may be more comprehensive by including these 2 risk factors, the measure effectively excludes over 50% of the facilities and 2/3rds of the records. This is a significant flaw. The incremental gain from including these 2 risk factors with a high prevalence of missing data on the C stat is very small (from 0.708 to 0.713). Excluding over 50% of hospitals in a measure intended for public reporting is problematic.

**Panel Member #4:** *N/A.*

**Panel Member #5:** Roughly half of hospitals performing TAVR are excluded.

**Panel Member #6:** *See Q #12.*

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

**Panel Member #6:** *The developers include variables such as race, sex, and age, which they argue are associated with SDS, but reject an explicit adjustment for social risk factors.*

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☒ No

**Panel Member #4:** *If I am not mistaken, one of the risk adjustors appears to be access site. If the procedure starts with a procedural pause, then this variable would be "intraop" and its inclusion may not be ideal. Perhaps "non-viable femoral access" or "severe preoperative femoral artery pathology" may be more informative and reflect a pre-existing comorbidity rather than incorrectly portray the variable as part of the procedure that is at the discretion of the surgeon.*

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)
☒ Yes ☐ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No

16e**. Assess the risk-adjustment approach**

**Panel Member #1:** Details on the development process of the model were not provided but referenced (Arnold SV 2018).

The challenge with the model selected is that it is based on a hierarchical model which needs to be re-estimated for each reporting period to assess the updated site-specific intercept. This requires adequate ongoing statistical support, as opposed to having a fixed set of coefficients. If my understanding is correct, this does not enable the provision of a patient level predicted score at admission to support clinical decision making. However, this is not a limitation that negatively impacts the measurement requirements for this submission, and I assume the pros/cons of this strategy were considered.

The section on testing utility of the 5 meter walk test and baseline KCCQ score under 2b4.3 could be moved to 2b3.

**Panel Member #2:** Development of the risk-adjustment model was clearly the focus of this measure development effort. Despite the inclusion of a host of clinical data elements the discrimination and calibration (in the validation sample) was not that much improved from an administrative data specification. One might wonder about the burden/benefit of the extensive data collection.

**Panel Member #3:** Hierarchical risk adjustment model. Hospital performance is reported using the hospital adjusted odds ratio. C statistic (0.70) and calibration graphs were acceptable in validation data set.

**Panel Member #5:** The method for risk adjustment was strong and well justified: **"**Case mix adjustment was implemented using a hierarchical logistic regression mode with site-specific random intercept parameters…. Because the purpose of these models is for risk adjustment of outcomes for site reporting, all covariates deemed clinically relevant were retained in these nonparsimonious models…. Covariates were selected a prior and were not removed on the basis of their statistical significance." All fine. The conceptual and empirical basis for not including social risk factor was strong. The discrimination in the validation data set was adequate (0.70) and the calibration was very good.

**Panel Member #6:** *Overall, the risk adjustment strategy is strong***.**

**Panel Member #7:** The risk adjustment model includes a relatively large number of factors that are not significantly associated with mortality. It is unclear why these factors were retained. The steward presents no evidence that a more parsimonious model was considered. The c-statistic of the model is modest (0.703), possibly suggesting that the set of risk factors omitted important clinical risk factors. Notably, no interaction effects are included in the risk adjustment model.

**For cost/resource use measures ONLY:**

17. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

### VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☒ **Data element** ☐ **Both**

    **Panel Member #3:** Performed data element validity testing. Auditors re-abstracted medical record (365 records) and compared agreement of re-abstracted data elements with registry data using a prevalence-adjusted kappa statistic (PABAK). The level of agreement was almost perfect for most of the data elements.

    **Panel Member #6:** *Although the developers say that the testing is at the data element level, their "empirical validity" testing relates to the goodness-of-fit of the risk adjustment model, which is at the measure score level.*

20. **Method of establishing validity of the measure score:**

    ☐ **Face validity**

    ☒ **Empirical validity testing of the measure score**

    ☒ **N/A (score-level testing not conducted)**

21. **Assess the method(s) for establishing validity**

    **Submission document: Testing attachment, section 2b2.2**

    **Panel Member #1:** Record eligibility assessment seemed appropriate and correctly done.

    Data element validity was tested only on the same 6 elements tested for percent agreement (reliability). As noted for the reliability data element testing, this is a very partial test as the large majority of data elements were not tested. No justification for selecting this sub-group of elements for testing was provided.

    Empirical validity was marked as being conducted on the score level. I was not clear on whether that was done or not. It seemed that the empirical validity testing was done on the patient level, graphically comparing observed and risk adjustment probabilities for the development and validation samples (section 2b3.8), and between several sub-groups (supplement figure).

    This looks to me more of a cross-validation test for the risk-adjustment model, not so much an empirical testing of the measure in comparison to another measure.

    However, the supplement figure establishes known groups validity (although not labeled as such) by demonstrating differences in observed and expected scores by patient groups for age, sex, ejection fraction, NYHA, and prior aortic pressure. If these differences follow expected clinical patterns, this may be interpreted as sufficient evidence to support the measure's validity.

    Additional information was referenced to section 2b4 but could not be identified within that section (performance differences). I assume this is a typo and 2b3 was the correct reference.

**Panel Member #2:** No concern. Trained auditors re-abstracted data elements from the medical record as a gold standard for comparison with registry submitted data. The PABAK statistic (a prevalence-adjusted and bias-adjusted KAPPA statistic) demonstrated data element validity.

**Panel Member #3:** Validity of risk adjustment model, as assessed using the C statistic and calibration graphs, implies predictive validity of the measure. This is acceptable. Discrimination and calibration were acceptable.

**Panel Member #4:** *Discrimination (c 0.703) and calibration (plot) in validation cohort.*

**Panel Member #5:** Method 1 was Records Eligibility Assessment (REA): checking to see if cases that qualified for the registry were represented in the registry. Method 2 data element validity: checking that the abstractions of a trained auditor of 6 "preselected data elements" matched the registry entries. Metrics were accuracy (presumably absolute agreement?) and PABAK statistics (Prevalence-adjusted bias-adjusted kappa). The mapping of value ranges to interpretations is not cited. Overall, these are reasonable methods to assess data element validity.

**Panel Member #6:** *The testing report is very confusing on this issue, and it is hard for me to see why the results they provide assess validity. Rather, they seem to address (a) reliability and (b) the goodness-of-fit of the risk adjustment model, neither of which addresses validity. The empirical testing also does not address all of the data elements.*

**Panel Member #7:** The steward assessed whether assessed cases were in agreement with billing code lists, and whether data elements were in agreement

22. **Assess the results(s) for establishing validity**

**Submission document: Testing attachment, section 2b2.3**

**Panel Member #1:** No concerns about results.

**Panel Member #2:** As the developer notes some of the PABAK statistics are lower than desired (0.77-0.82) although still demonstrating substantial agreement. Specifically follow-up status is particularly problematic given that the outcome of interest is 30-day mortality. Also of concern is the small number of cases (N=8) with a data of death and the low PABAK statistics (0.5) that demonstrates moderate agreement. An over-sampling of records with a data of death might be preferred.

**Panel Member #3:** Validity of risk adjustment model, as assessed using the C statistic and calibration graphs, implies predictive validity of the measure. This is acceptable. Discrimination and calibration were acceptable.

**Panel Member #4:** *Adequate. I suspect there may be some discounting of the c-statistic given this population's homogeneity, however, I cannot demonstrate this nor do the materials speak directly to this possibility.*

**Panel Member #5:** Results for Method 1 (REA) were very good. Eligible cases appear to be represented in the registry.

Results for Method 2 (data element validity) were mixed. The submission states that Agreement <85% indicated that the validity of the data element needs improvement. One of the 6 variables (FU date of death) had agreement of 75% and 2 others identified some sites with low agreement (percentiles are from the hospital-level distribution of agreement?) As stated "The agreement rates of Follow-up Status were lower than expected. Results were left skewed with one site having 0.0% agreement. Thirty-four of the 44 mismatches were a result of no 30-day follow-up status being submitted to the registry despite documentation that supported a 30-day follow-up status of alive or deceased status. The other ten had submitted a follow-up status of alive or dead and there was no documentation present for the auditor to validate the answer during review". Agreement for follow-up date of death was lower and more varied. Areas for improvement were noted for the 5 meter walk test element. Unclear why other data elements were not evaluated. Given the variable validity of the items that were

checked (including the outcome), and the unknown reliability of other casemix variables, I am unsure if the current evidence supports implementation.

**Panel Member #6:** *Because the methods do not seem appropriate, I don't think that the results are informative.*

**Panel Member #7:** Only one of 365 cases was flagged for omission and potential eligibility. Regarding data elements, agreement was generally very high.

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    **Submission document:** Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

    ☒ **Not applicable** (score-level testing was not performed)

    **Panel Member #1:** As noted above, this is my understanding given the information provided.

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

    **Submission document:** *Testing attachment, section 2b1.*

    ☒ **Yes**

    ☒ **No**

    **Panel Member #1:** Not all data elements were tested, only a small amount of elements were assessed.

    ☒ **Not applicable** (data element testing was not performed)

    **Panel Member #2:** No, only limited data element validity was performed. However, it would have been more resource intensive to examine the data element validity for all 41 data elements.

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

    ☒ **Low** (NOTE: Should rate LOW if you believe that there <u>are</u> threats to validity and/or relevant threats to validity were <u>not assessed OR</u> if testing methods/results are not adequate)

    ☒ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level <u>is required</u>; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

    **Panel Member #1:** The insufficient rate is due to clarifications needed about the empirical validity testing. Was it done on the data element level or score level? Is 'empirical validity' the correct term? As noted above, it seems more of a cross-validation method for the risk-adjustment modeling, incorporated with know-groups validity for several patient sub-groups (calibration plots shown in the supplementary figure). Additional clarifications are needed.

    **Panel Member #2:** Given the validity testing results the rating is borderline low although there is a substantial infrastructure in place to address data element validity concerns. There is a statement that "additional empirical validity testing was unnecessary because the importance of a low mortality rate is undisputed and it can be

measured directly" which seems to imply that score level validity testing is unnecessary because a mortality outcome is valid on its face. The statement misunderstands the nature and purpose of score level validity testing.

**Panel Member #3:** Validity of risk adjustment model, as assessed using the C statistic and calibration graphs, implies predictive validity of the measure. This is acceptable. Discrimination and calibration were acceptable.

**Panel Member #4:** *Adequate discrimination and calibration.*

**Panel Member #5:** If it is expected that the reliability and validity of all case mix variable are assessed, then the evidence presented here is incomplete. That aside, the analysis of reliability was not problematic from a methodological perspective. The analysis of Record eligibility assessment was strong and convincing. The methodology for data element validity was somewhat underdescribed, but probably appropriate. The analyses on the 6 data elements reveals generally reasonable validity but also somewhat concerning site-level variation and quality improvement opportunities for key data elements. The distribution of the measure scores (odds ratios) across eligible sites shows good variability.

**Panel Member #6:** *The methods seem to be totally inappropriate.*

**Panel Member #7:** Documentation is sparse, but analysis appears to indicate a potentially low validity of follow-up status and date of death information. This is a direct threat to the measure, as death is the outcome of interest.

### FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

### ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

    **Panel Member #1:** I believe the SMP should discuss these concerns before forwarding them to the standing committee.

    **Panel Member #2:** The risk adjustment model is well-specified but the developer needs some technical assistance in planning for more comprehensive and compelling reliability and validity testing in measure maintenance.

## Developer Submission

# 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**

3534_NQF_evidence_attachment_TAVR30RAM_11_6_2019.docx

**1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?**

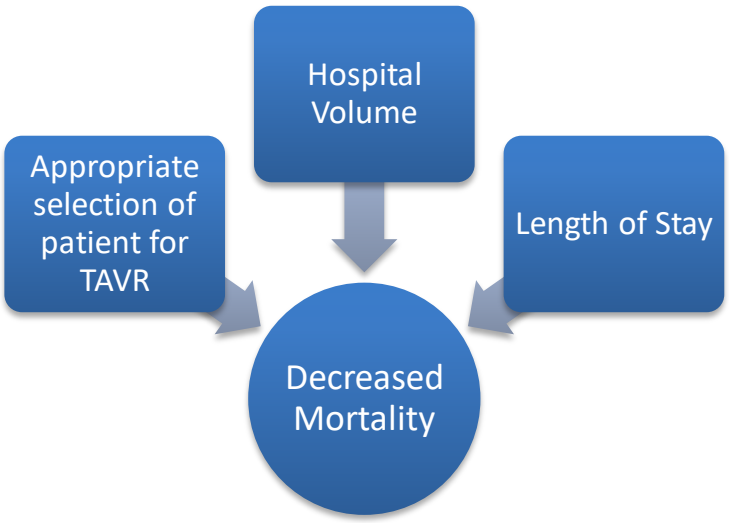
Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

### 1a. Evidence (subcriterion 1a)

**Measure Number** (*if previously endorsed*)**:** 3534
**Measure Title**: 30 Day All-cause Risk Standardized Mortality Odds Ratio following Transcatheter Aortic Valve Replacement (TAVR)
**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**: 11/8/2019

**1a.1.This is a measure of**: (*should be consistent with type of measure entered in De.1*)
Outcome
☒ Outcome: 30 Day All-cause Mortality following Transcatheter Aortic Valve Replacement (TAVR)
☐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:
☐ 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.




This measure examines hospital-level 30 day mortality rates following TAVR, with the overriding goal to reduce 30-day mortality rates to best-in-class. The expectation is that providing this information to hospitals, coupled with public reporting of hospitals' results, will drive internal hospital quality improvement efforts to focus efforts on reducing TAVR mortality. Of note, the measure includes in-hospital deaths and deaths occurring after hospital discharge up to 30 days post procedure. This perspective may motivate hospitals to look for opportunities not only within the organization, but to better coordinate the transition of care from the inpatient to the outpatient arena.

**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.)

Not applicable

****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.**

Several studies demonstrate that there are at least two factors within a hospital's control by which 30-day mortality rates can be impacted – appropriate patient selection and a hospital's overall volume of TAVR.

Arnold[1] examined whether a worse pre-procedure patient health status, as assessed by the KCCQ, was associated with greater long-term mortality after TAVR. An analysis of TVT Registry data from 2011 to 2014 with 304 hospitals found that patients with worse health status were more likely to be female, had more comorbidities, and higher STS risk scores for mortality. Compared with those with good health status before TAVR and after adjusting for a broad range of baseline covariates, patients with very poor health status had a 2-fold increased hazard of death over the first year after TAVR whereas those with poor and fair health status had intermediate outcomes. These results demonstrate that appropriate patient selection and mortality risk assessments for patients considering TAVR can directly influence mortality.

Two other studies[2,3] examined whether there is a relationship between volume of TAVRs and outcomes, including mortality. An analysis of TVT registry data from 2011 to 2015 with 395 hospitals found that higher volumes were

associated with in-hospital mortality as well as vascular complications and stroke, particularly with those hospitals with less than 100 procedures (Carroll, 2017). Vemulapalli (2019) summarized the volume/outcome experience of the TVT Registry from 2015 to 2017 with 113,662 TAVR procedures were performed at 555 hospitals by 2960 operators. The authors noted mortality at 30 days was higher and more variable at hospitals with a low procedural volume than at hospitals with a high procedural volume. These results further validate the relationship between increased site volume with lower mortality rates.

[1]Arnold, SV, Spertus, JA, Vemulapalli, S, et al. Association of Patient Reported Health Status With Long-Term Mortality After Transcatheter Aortic Valve Replacement. A Report from the STS/ACC TVT Registry. Circulation Cardiovascular Interventions. 2015; 8: e002875.

[2]Carroll JD, Vemulapalli S, Dai D. Procedural experience for transcatheter aortic valve replacement and relation to outcomes. JACC. 2017;70:29–41.

[3]Vemulapalli, S., et al. Procedural Volume and Outcomes for Transcatheter Aortic Valve Replacement. NEJM, April 2, 2019, p1-11.

**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:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | |
| Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR. | |

| | |
|---|---|
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | |
| Provide all other grades and definitions from the evidence grading system | |
| Grade assigned to the **recommendation** with definition of the grade | |
| Provide all other grades and definitions from the recommendation grading system | |
| Body of evidence:<br>• Quantity – how many studies?<br>• 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.*

**1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure.** A list of references without a summary is not acceptable.

**1a.4.2 What process was used to identify the evidence?**

**1a.4.3. Provide the citation(s) for the evidence.**

## 1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

**1b.1. Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

*If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.*

This measure will describe hospital-level 30-day mortality rates following TAVR, with the overriding goal to reduce 30-day mortality rates. The expectation is that providing this information to hospitals, coupled with public reporting of hospitals' results, will drive internal hospital quality improvement efforts to focus efforts on reducing TAVR mortality. Of note, the measure includes in-hospital deaths and deaths occurring after hospital discharge up to 30 days post procedure. This perspective may motivate hospitals to look for opportunities not only within the organization, but to better coordinate the transition of care from the inpatient to the outpatient arena.

**1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

For this NQF submission, we use TVT registry data from two partially overlapping 3-year time periods:

- June 2013 – May 2016 (the original development and validation data)
  - 21,661 TAVR patients from 188 TVT hospitals
- April 2015 – March 2018 (most recent 3-year data available)
  - 49,182 TAVR patients from 265 TVT hospitals

The distribution of hospital-specific odds ratio estimates are:

June 2013 – May 2016:

Mean: 1.00

Standard Deviation: 0.02

Minimum: 0.92

Maximum: 1.07

25th: 0.99

75th: 1.01

10th: 0.97

20th: 0.98

30th: 0.99

40th: 1.00

50th: 1.00

60th 1.01

70th: 1.01

80th: 1.02

90th: 1.03

April 2015 – March 2018:

Mean: 1.01

Standard Deviation: 0.10

Minimum: 0.81

Maximum: 1.40

25th: 0.95

75th: 1.06

10th: 0.89

20th: 0.94

30th: 0.96

40th: 0.98

50th: 1.00

60th 1.02

70th: 1.04

80th: 1.07

90th: 1.13

**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.**

Death is the indicator that has been most widely used to evaluate the quality of cardiac procedures and is arguably the most important adverse outcome measure from the perspective of the patient.

Arnold (2) summarized that the risk adjustment model facilitates objective comparisons of outcomes across sites by accounting for differences in case mix. Importantly, these models included patient-reported health status and gait speed at baseline, two factors known to be prognostically important beyond traditional demographic and clinical factors and never used in models previously (1,3).

Vemulapalli (7) summarized the volume/outcome experience of the TVT Registry from 2015 to 2017. 113,662 TAVR procedures were performed at 555 hospitals by 2960 operators. Mortality at 30 days was higher and more variable at hospitals with a low procedural volume than at hospitals with a high procedural volume. This validated Carroll's (4) findings that increasing site volume is associated with lower mortality rates.

Citations

1Alfredsson J, Stebbins A, Brennan JM, et al. Gait speed predicts 30-day mortality after transcatheter aortic valve replacement: results from the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. Circulation 2016;133:1351–9.

2Arnold, S.V. et al. Measures in the Risk Adjustment of 30-Day Mortality After Transcatheter Aortic Valve Replacement: A Report From the STS/ACC TVT Registry JACC: Cardiovascular Interventions Volume 11, Issue 6, 26 March 2018, Pages 581-589

3Arnold SV, Spertus JA, Vemulapalli S, et al. Association of patient-reported health status with long-term mortality after transcatheter aortic valve replacement: report from the STS/ACC TVT Registry. Circ Cardiovasc Interv 2015;8.

4Carroll, J.D., et al. Procedure Experience for TAVR and Relation to Outcomes, The STS/ACC TVT Registry. JACC, Vol 70, #1, 2017. Page 29-41.

5Grover FL, Vemulapalli S, Carroll JD, et al. 2016 Annual report of the STS/ACC Transcatheter Valve Therapy Registry. J Am Coll Cardiol 2017; 69: 1215-30.

6O'Brien, Sean, et al. Variation in Hospital Risk – Adjusted Mortality Rates Following TAVR in the U.S. A Report from the STS/ACC TVT Registry. Circ Outcomes. Sept 2016

7Vemulapalli, S., et al. Procedural Volume and Outcomes for Transcatheter Aortic Valve Replacement. NEJM, April 2, 2019, p1-11.

**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.*

In order to explore disparities, we modified the measure´s hierarchical model to include indicator variables for black race, other non-white race, Hispanic ethnicity, and participation in Medicaid. We performed this analysis using data from June 2013 to May 2016 (21,661 patients from 188 hospitals) and using data from April 2015 to March 2018 (49, 182 patients from 264 hospitals). In order to accommodate these variables, we removed an existing related variable that was defined as "non-white race or Hispanic ethnicity". Results are summarized by in the form of odds ratios below. For each variable in each time period, the 95% confidence interval around the odds ratio overlaps with the null value of 1.0. This implies that there was no statistically significant association between these variables and 30-day mortality after adjusting for other factors in the hierarchical model ($p>0.05$ for each variable below).

June 2013 – May 2016

Medicaid: 0.93 (0.69 - 1.24)

Black race (versus white): 0.73 (0.47 - 1.14)

Other non-white race (versus white): 0.68 (0.36 - 1.30)

Hispanic ethnicity: 1.23 (0.82 - 1.84)

April 2015 – March 2018

Medicaid: 1.05 (0.83 - 1.32)

Black race (versus white): 0.91 (0.67 - 1.23)

Other non-white race (versus white): 0.85 (0.54 - 1.34)

Hispanic ethnicity: 0.82 (0.59 - 1.16)

**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.)*

https://www.ncdr.com/WebNCDR/docs/default-source/tvt-public-page-documents/coderdatadictionary_pdf-(1).pdf?sfvrsn=2

**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:** TAVR_S.2b_attachment-637092425369121221.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 (not maintenance of endorsement)

**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 of this measure is all-cause death within 30 days following a transcatheter aortic valve replacement (TAVR).

**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).*

NUMERATOR:

1. Discharge status of expired or

2. Follow-up status=deceased and date difference between index procedure and death date is <=30 or

3. 30-day follow-up status=deceased, death date is missing, and difference between index procedure and follow-up assessment date is <=75 days. *

*Notes: The <=75 day follow-up assessment timeframe was identified to be a clinically reasonable surrogate to capture a 30 day death if 30 day follow-up date of death was missing (this occurred in 0.9% of deceased records from January 2015 to December 2017). Sometimes a status of "deceased" is known and documented but the exact date of death is not available.

In addition, we validated the accuracy of 30-day mortality in the TVT Registry by comparing Registry data linked CMS claims data from 2012-2015. Across 3.5 years, 99.6% of the 29,247 patient records had no discrepancy.

**S.6. Denominator Statement** *(Brief, narrative description of the target population being measured)*

The target population for the outcome is for individuals who have undergone transcatheter aortic valve replacement.

For development, reassessment and reporting of this measure, we use site reported data from the STS/ACC TVT Registry.

**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).*

Measure Eligibility and Population Definition

1) Eligibility at the hospital level:

a) Acceptable "Data Quality Report" data submissions for each quarter in the reporting period.

b) Hospitals must have >=90% completeness of the following items for all patient records in the rolling 3-year reporting period to receive feedback on the measure:

i) Computed baseline Kansas City Cardiomyopathy Questionnaire (a key risk model covariate) AND

ii) Baseline 5-meter walk test (a key model covariate), AND

iii) 30-day follow-up status =alive or dead as defined above (the outcome variable)

2) Eligibility at the patient level: Hospitalization for first-time TAVR procedure

**S.8. Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

1) Hospitals need to meet eligibility criteria to be included in the measure.

2) Patients are excluded if:

a) They did not have a first-time TAVR in the episode of care (admission),

b) The TAVR was subsequent to another procedure in the Registry (other TAVR, Mitral Leaflet Clip and/or TMVR) during that admission.

c) The patient is readmitted for a repeat TAVR (re-admission) and the initial TAVR was performed during the rolling 3-year timeframe for the measure.

d) 30-day mortality status missing.

**S.9. Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

1) Hospital ineligibility:

a) Unacceptable data quality report submissions for all quarters of the reporting time-period.

b) Hospitals who have less than 90% of patient records with respect to ANY of the following assessments in the rolling 3-year reporting period:

i) Computed baseline Kansas City Cardiomyopathy Questionnaire (a key risk model covariate) OR

ii) Baseline 5 meter walk test (a key model covariate), OR

iii) 30 day follow-up status =alive or dead as defined above (the outcome variable)

2) Patient Ineligibility:

a) They did not have a first-time TAVR in the episode of care (admission),

b) The TAVR was subsequent to another procedure in the Registry (other TAVR, Mitral Leaflet Clip and/or TMVR) during that admission.

c) The patient is readmitted for a repeat TAVR (re-admission) and the initial TAVR was performed during the rolling 3-year timeframe for the measure.

d) 30-day mortality status is missing.

**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.)*

This measure will not be stratified.

**S.11. Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)

Statistical risk model

If other:

**S.12. Type of score:**

Ratio

If other:

**S.13. Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Lower score

**S.14. Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

The measure score is calculated based on the following steps:

1) Patient cohort is identified based on inclusion criteria (see questions S.7-S.11)

2) Data elements for risk adjusted are collected using the first collected value, as identified below;

3) Outcome is ascertained (see S.5)

4) Measure score is calculated with aggregated data across all included sites as described below. Risk adjustment variables include:

1. Age
2. Body surface area (BSA)
3. Sex
4. Race/ethnicity
5. Estimated glomerular filtration rate (eGFR), which quantifies kidney function
6. Hemodialysis for end-stage renal disease
7. Left ventricular ejection fraction (LVEF)
8. Hemoglobin
9. Platelet count
10. Procedure date
11. Left main coronary artery stenosis = 50%
12. Proximal left anterior descending coronary artery stenosis = 70%
13. Prior myocardial infarction
14. Endocarditis
15. Gait speed (via the 5-meter walk test which assesses frailty)
16. Baseline Kansas City Cardiomyopathy Questionnaire-12 (KCCQ-12, a measure of heart-failure specific health status)
17. Peripheral artery disease
18. Current/recent smoker
19. Diabetes
20. Atrial fibrillation/flutter
21. Conduction defect
22. Chronic lung disease
23. Home oxygen
24. “Hostile” chest
25. Porcelain (severely concentrically calcified) aorta
26. Access site
27. Pacemaker
28. Previous implantable cardioverter defibrillator
29. Prior percutaneous coronary intervention
30. Prior coronary artery bypass surgery
31. # prior cardiac operations

32. Prior aortic valve surgery/procedure

33. Prior other valve procedure surgery/procedure (mitral, tricuspid, pulmonic)

34. Aortic valve disease etiology

35. Aortic valve morphology

36. Aortic insufficiency (moderate or severe)

37. Mitral insufficiency (moderate or severe)

38. Tricuspid insufficiency (moderate or severe)

39. Acuity status (defined by a combination of procedure status, prior cardiac arrest w/in 24 hours, need for pre-procedure inotropic medications, and use of mechanical assist device)

40. Carotid stenosis

41. Prior transient ischemic attack or stroke

Case mix adjustment is implemented using a hierarchical logistic regression model with the above covariates and a site-specific random intercept. The main summary measure of a hospital´s risk-adjusted outcomes performance is the hospital´s estimated odds ratio, which compares the predicted odds of death of the patient population at a hospital if TAVR is performed by the hospital of interest to the predicted odds of death if TAVR were performed by an average hospital. An odds ratio greater than 1 implies higher than expected mortality and an odds ratio less than 1 implies lower than expected mortality. Each hospital´s estimated odds ratio is reported along with an approximate 95% empirical Bayes interval around the estimated odds ratio.

Definition of Measure Score Calculation - Odds ratio: a parameter reflecting the association between risk factors and an outcome.

The Risk Standardized Odds Ratio is calculated as the odds that an outcome (e.g. 30-day mortality) will occur for patients treated at your facility compared to the "odds" that outcome will occur for patients with identical risk factors if treated by a hypothetical (average) hospital.

It conceptually allows for a comparison of a particular hospital's performance given its case mix to an average hospital's performance with the same case mix. Thus, a lower odds ratio implies lower-than-expected mortality (better quality) and a higher ratio implies higher-than-expected mortality (worse quality). To assess hospital performance in any reporting period, we re-estimate the model coefficients using the years of data in that period.

References:

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22 (2): 206-226.

Arnold, S.V. et al. Measures in the Risk Adjustment of 30-Day Mortality After Transcatheter Aortic Valve Replacement: A Report From the Society of Thoracic Surgeons/American College of Cardiology TVT Registry JACC: Cardiovascular Interventions Volume 11, Issue 6, 26 March 2018, Pages 581-589

**S.15. Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A. This measure is not based on a sample or survey. Data from all hospitals and all TAVR procedures would be included in the process of re-estimating model variables. For public reporting, minimum sample size has not been determined.

**S.16. Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

N/A. This measure is not based on a sample or survey.

**S.17. Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

*If other, please describe in S.18.*

Registry Data

**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.

STS/ACC TVT Registry

**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)*

Available at measure-specific web page URL identified in S.1

**S.20. Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility

**S.21. Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Inpatient/Hospital

If other:

**S.22. COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

N/A

**2. Validity – See attached Measure Testing Submission Form**

TAVR_nqf_testing_attachment_7.31.2019_updated_11_7_19.docx

**2.1 For maintenance of endorsement**

*Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

No

**2.2 For maintenance of endorsement**

*Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

No

**2.3 For maintenance of endorsement**

*Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.*

## Measure Testing (subcriteria 2a2, 2b1-2b6)

**Measure Title:** 30 day All-cause Risk Standardized Mortality Odds Ratio following Transcatheter Aortic Valve Replacement (TAVR)
**Date of Submission:** July 31, 2019, updated November 5, 2019
**Type of Measure:**

| | |
|---|---|
| ☒ Outcome (*including PRO-PM*) | ☐ Composite – ***STOP – use composite testing form*** |
| ☐ Intermediate Clinical Outcome | ☐ Cost/resource |
| ☐ Process *(including Appropriate Use)* | ☐ Efficiency |
| ☐ Structure | |

**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: Click here to describe | ☐ other: Click here to describe |

**1.2. If an existing dataset was used, identify the specific dataset** (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

The TVT Registry was launched in 2011 as a joint initiative of the Society of Thoracic Surgeons and the American College of Cardiology. It now includes more than 450 clinical sites across the United States. Hospitals are required to participate in the registry by Medicare to obtain reimbursement for TAVR. Accordingly, data on nearly all TAVR procedures performed outside of clinical trials in the United States are captured in the registry. To promote quality improvement efforts both locally and nationally, participating centers receive quarterly reports comparing each center's case mix, practice patterns, and outcomes to the national experience.

**1.3. What are the dates of the data used in testing**?

Data element reliability and validity testing: January 1, 2016 – December 31, 2017

All other testing:

The measure was developed and tested using TVT Registry data from the 3-year period June 2013 – May 2016. For this NQF submission, we use data from two partially overlapping 3-year periods:

- June 2013 – May 2016 (the original development and validation data)
- April 2015 – March 2018 (most recent 3-year data available)

TVT Registry data was linked CMS claims data from 2012-2015. This data linking was used to validate the reporting the accuracy of death in follow-up assessments.

**1.4. What levels of analysis were tested**? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

| **Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)** | **Measure Tested at Level of:** |
|---|---|
| ☐ individual clinician | ☐ individual clinician |
| ☐ group/practice | ☐ group/practice |
| ☒ hospital/facility/agency | ☒ hospital/facility/agency |
| ☐ health plan | ☐ health plan |
| ☐ other: Click here to describe | ☐ other: Click here to describe |

**1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Data element reliability testing:
Inter-rater reliability (IRR) audits were conducted on four (4) randomly selected facilities from a nationwide audit for a total of 40 randomly selected records each year.

Data element validity testing:
Forty (40) hospitals that report to the STS/ACC TVT Registry were randomly selected nationwide to assess the data accuracy and records eligibility for all procedures (TAVR, mitral repair and mitral replacement) in an audit for the period between 01/01/16 and 12/31/16, as well as between 01/01/17 and 12/31/17.

In addition to data validity testing, six (6) of these hospitals were also asked to participate in the "records eligibility assessment" (to verify records submitted to the registry aligned with CPT, ICD-9, and ICD-10 codes of procedures performed in both 2016 and 2017).

All other testing:

**Table 1**

| **June 2013 – May 2016** | **April 2015 – March 2018** |
|---|---|
| 188 hospitals | 264 hospitals |

**1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Data element reliability testing:
Inter-rater reliability (IRR) audits were conducted on randomly selected records and facilities as follows:

- 2017: 10 sites across 10 auditors with a total of 105 cases (55 baseline and 32 30-day)
- 2016: 4 sites across two auditors with a total of 70 cases (10 baseline and 10 30-day).

Data element validity testing:
2017:

- 1,000 procedures at baseline (admission to discharge) and 30 day follow-up (includes 910 TAVR and 90 mitral repair or replacement procedures)
- 380 records for the records eligibility assessment

2016:

- 400 procedures at baseline (admission to discharge) and 30-day follow-up (includes 356 TAVR and 44 mitral repair or replacement procedures)
- 343 records for the records eligibility assessment

All other testing:
For the original development and testing sample, we selected TVT records from all patients undergoing TAVR at a TVT hospital during June 1, 2013 to May 31, 2016 (N = 60,770 records, 450 hospitals). Only the first TAVR record per patient was included. The analysis was restricted to hospitals with ≥90% complete non-missing data for key study variables including 30-day status, baseline KCCQ-12 score, and baseline gait speed (N = 22,506 records, 188 hospitals). Patients at these sites who had missing data for 30-day status were excluded (N = 845). The final cohort was 21,661 patients from 188 hospitals. Patient characteristics are summarized in Table 2.

**Table 2**

| **Risk Factors** | **All patients (N=21661)** | **All patients % (N=21661)** |
|---|---|---|
| Age Categories | | |
| Age < 75 yrs | 4474/21661 | 20.7% |
| Age 75-84 yrs | 7973/21661 | 36.8% |
| Age >=85 yrs | 9214/21661 | 42.5% |
| BSA Categories | | |
| BSA < 1.80 m2 | 8957/21650 | 41.4% |
| BSA 1.80-2.19 m2 | 10610/21650 | 49.0% |
| BSA >=2.20 m2 | 2083/21650 | 9.6% |
| Female | 10486/21661 | 48.4% |
| Race, non-white or Hispanic | 1369/21661 | 6.3% |
| Renal Function Categories | | |
| no dialysis GFR >= 60 mL/min/1.73m2 | 10975/20826 | 52.7% |
| no dialysis GFR 30-59 mL/min/1.73m2 | 8759/20826 | 42.1% |
| no dialysis GFR <30 mL/min/1.73m2 | 1092/20826 | 5.2% |
| current dialysis | 805/21649 | 3.7% |
| Ejection Fraction | | |
| LVEF < 35% | 2371/21556 | 11.0% |
| LVEF 35-54% | 5321/21556 | 24.7% |
| LVEF >= 55% | 13864/21556 | 64.3% |
| Hemoglobin, < 10 mm/dL | 3115/21638 | 14.4% |

| Risk Factors | All patients (N=21661) | All patients % (N=21661) |
|---|---|---|
| Platelet, < 100 uL | 873/21603 | 4.0% |
| Left main disease >= 50% | 2155/21526 | 10.0% |
| Proximal LAD >=70% | 4282/21517 | 19.9% |
| Prior MI | 5426/21628 | 25.1% |
| Endocarditis | 181/21630 | 0.8% |
| Prior TIA or stroke | 4172/21661 | 19.3% |
| Carotid Stenosis | 5074/21661 | 23.4% |
| Prior PAD | 6637/21661 | 30.6% |
| Smoker | 1247/21661 | 5.8% |
| Diabetes | 8108/21661 | 37.4% |
| Atrial fibrillation or flutter | 8971/21661 | 41.4% |
| Conduction Defect | 7901/21661 | 36.5% |
| Severe chronic lung disease | 3028/21661 | 14.0% |
| Home oxygen use | 2664/21661 | 12.3% |
| Hostile chest | 1509/21661 | 7.0% |
| Porcelain aorta | 1266/21661 | 5.8% |
| Access site, non-femoral | 4655/21661 | 21.5% |
| Previous ICD | 845/21661 | 3.9% |
| Prior PCI | 7542/21661 | 34.8% |
| Prior CABG | 5991/21661 | 27.7% |
| Prior aortic procedure | 2776/21661 | 12.8% |
| Prior non-aortic procedure | 507/21661 | 2.3% |
| Aortic etiology, degenerative | 20702/21661 | 95.6% |
| Valve morphology, tricuspid | 20513/21661 | 94.7% |
| Aortic insufficiency, at least moderate | 4395/21661 | 20.3% |
| Mitral insufficiency, at least moderate | 6426/21661 | 29.7% |
| Tricuspid insufficiency, at least moderate | 5291/21661 | 24.4% |
| Acuity category 2 | 1652/21661 | 7.6% |
| Acuity category 3 | 626/21661 | 2.9% |
| Acuity category 4 | 86/21661 | 0.4% |
| Unable to walk | 2551/21661 | 11.8% |
| Speed by quartiles | | |
| Speed <0.417 | 5488/21661 | 25.3% |
| Speed 0.417-0.625 | 6183/21661 | 28.5% |
| Speed 0.625-0.789 | 4361/21661 | 20.1% |
| Speed >0.789 | 5629/21661 | 26.0% |
| KCCQ overall by quartiles | | |
| KCCQ <23.96 | 5421/21661 | 25.0% |
| KCCQ 23.96 -40.10 | 5664/21661 | 26.1% |
| KCCQ 40.10-58.33 | 4970/21661 | 22.9% |
| KCCQ >58.33 | 5606/21661 | 25.9% |

We also performed analyses using data from the most recent 3-year period for which data was available (April 1, 2015 to March 31, 2018). This cohort included 98,364 TAVR procedures performed at 264 hospitals.

**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**.

Data element reliability and validity testing used a randomly selected sample of 40 hospitals participating in the TVT Registry between January and December 2016 and a randomly selected sample of 50 hospitals between January and December 2017.

All other testing included all hospitals and patients that met the inclusion criteria in the same registry between June 2013 – May 2016.

**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.

Whether outcomes measures, and the public reporting and reimbursement programs based on them, should consider socioeconomic (SES) or sociodemographic (SDS) factors (e.g., race, ethnicity, education, income, payer [e.g., Medicare-Medicaid dual eligible status]) is a topic of intense health policy debate [1]. Some argue that in the absence of adjustment for these variables, the outcomes of hospitals that care for a disproportionate percentage of low SES patients will be unfairly disadvantaged, perhaps leading to financial or reputational penalties. Opponents argue that inclusion of SES factors in risk models may "adjust away" disparities in quality of care, and they advocate the use of stratified analyses instead. They also note that readily available SES factors have often not demonstrated significant impact on outcomes. As part of an NQF pilot project, STS specifically studied dual eligible status in the STS readmission measure [2] and found minimal impact. Finally, even proponents of inclusion of SES in risk models agree that these factors make more sense intuitively for some outcomes (e.g., readmission) than others (hospital mortality, complications)—that is, they are context-specific [2,3].

In identifying a risk adjustment approach for this measure, and in keeping with the general approach taken for the current risk models by the Society for Thoracic Surgeons [3], we chose to avoid the more philosophical and downstream health policy implications of SES adjustment and based our modeling decisions on empirical findings and consideration of the model's primary intended purpose--to adjust for case mix. Conceptually, our goal was to adjust for all preoperative factors that are independently and significantly associated with outcomes and that vary across TVT participants. For example, race and ethnicity will continue to be in our risk models as it has been previously, but not conceptually as a SES indicator. Race has an empirical association with outcomes and has the potential to confound the interpretation of a hospital's outcomes, although we do not know the underlying mechanism (e.g., genetic factors, differential effectiveness of certain medications, rates of certain associated diseases such as diabetes and hypertension).

For purposes of this NQF submission, we did perform analyses of race, ethnicity, and Medicaid status (see submission form, section 1.b.4). Findings in this analysis implies that there was no statistically significant association between these variables and 30-day mortality after adjusting for other factors in the hierarchical model.

1. National Academies of Sciences E, and Medicine. Accounting for social risk factors in medicare payment. Washington, DC: The National Academies Press; 2017.

2. Shahian DM, He X, O'Brien SM et al. Development of a clinical registry-based 30-day readmission measure for coronary artery bypass grafting surgery. Circulation 2014;130(5):399-409.

3. Shahian DM, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1 – Background, Design Considerations, and Model Development. Ann Thorac Surg. 2018 May;105(5):1411-1418.

**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*)

Two trained auditors performed inter-rater reliability (IRR), performing a visual inspection of the medical record for the sample cases (each reviewed the records in the sample) to abstract necessary data. IRR assessment was performed on baseline and 30-day follow-up cases.

**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*)

Forty (40) records were reviewed in 2016 and fifty (50) records in 2017 to ensure that the auditors were abstracting the data consistently. Agreement rates and PABAK scores were calculated in Table 3.

**Table 3**

| Data Element | Proc. Type | Matches | Universe | Agreement Rate (IRRA) | PABAK | | |
|---|---|---|---|---|---|---|---|
| | | | | | Score | Lower 95% CI | Upper 95% CI |
| Birth Date (DOB) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Sex (Sex) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Permanent Pacemaker (Pacemaker) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Prior PCI (PriorPCI) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Prior CABG (PriorCABG) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Prior Aortic Valve Procedure (PriorAorticValve) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Prior Stroke (PriorStroke) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| TIA (CVDTIA) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Peripheral Arterial Disease (PriorPAD) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Diabetes Mellitus (Diabetes) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |

| Data Element | Proc. Type | Matches | Universe | Agreement Rate (IRRA) | PABAK | | |
|---|---|---|---|---|---|---|---|
| | | | | | Score | Lower 95% CI | Upper 95% CI |
| Currently on Dialysis (CurrentDialysis) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Chronic Lung Disease (ChrLungD) | All | 52 | 55 | 94.5% | 0.891 | 0.73 | 1.00 |
| Home Oxygen (HMO2) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Hostile Chest (HostileChest) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Immunocompromise (ImmSupp) | All | 54 | 55 | 98.2% | 0.964 | 0.86 | 1.00 |
| Prior MI (PriorMI) | All | 54 | 55 | 98.2% | 0.964 | 0.86 | 1.00 |
| Porcelain Aorta (PorcelainAorta) | All | 54 | 55 | 98.2% | 0.964 | 0.86 | 1.00 |
| Atrial Fibrillation/Flutter (AFibFlutter) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Five Meter Walk Test Performed (FiveMWalkTest) | TAVR | 41 | 43 | 95.3% | 0.907 | 0.73 | 1.00 |
| KCCQ-12 Performed (KCCQ12_Performed) | All | 54 | 55 | 98.2% | 0.964 | 0.86 | 1.00 |
| Height (Height) | All | 51 | 55 | 92.7% | 0.855 | 0.67 | 1.00 |
| Weight (Weight) | All | 50 | 55 | 90.9% | 0.818 | 0.61 | 1.00 |
| Pre-Procedure Creatinine (PreProcCreat) | All | 53 | 55 | 96.4% | 0.927 | 0.79 | 1.00 |
| Left Ventricle Ejection Fraction (LVEF) | All | 51 | 55 | 92.7% | 0.855 | 0.67 | 1.00 |
| Procedure Start Date (TVTProcedureStartDate) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Valve Sheath Access Site (TVTAccessSite) | TAVR | 43 | 43 | 100.0% | 1.000 | 1.00 | 1.00 |
| Discharge Date (DCDate) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| Discharge Status (DCStatus) | All | 55 | 55 | 100.0% | 1.000 | 1.00 | 1.00 |
| **Baseline Overall Accuracy** | | 2315 | 2370 | **97.7%** | 0.954 | 0.94 | 0.97 |
| Follow-up Status (F_Status) | All | 32 | 32 | 100.0% | 1.000 | 1.00 | 1.00 |
| Follow-up Date of Death (F_DeathDate) | All | 0 | 0 | N/A | N/A | N/A | N/A |
| **30-Day Follow-up Overall Accuracy** | | 174 | 175 | **99.4%** | 0.989 | 0.96 | 1.00 |

**2a2.4 What is your interpretation of the results in terms of demonstrating reliability**? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The high agreement rate for all but one of the data elements indicates a good understanding of data

definitions and consistency between the auditors. It also provides assurance of the accuracy of the re-abstraction being performed among the auditing team over multiple years. While we are not able to assess all model variables because of competing regulatory requirements for post approval studies in the TVT Registry, 24 of the 41 model variables in addition to the 6 critical data elements for the numerator and denominator are provided. We are re-evaluating adding additional risk model variables in future years.

---

**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)*

Record eligibility assessment: For 2016 and 2017, six (6) hospitals who participated in an audit of the STS/ACC TVT Registry were asked to report all TAVR and Mitral cases performed at their facility during a specific period randomly selected and based on the specific billing codes for these procedures. This report was compared against the registry's list for the baseline records and for the same period to assess the records validity and verify cases are not missed.

Data element validity:
Forty (40) randomly selected hospitals in 2016 and fifty (50) in 2017 were chosen to participate in an audit of the STS/ACC TVT Registry. Sites with a minimum of 10 baseline records during the audit period (01/01/2016 – 12/31/2016 and 01/01/2017 – 12/31/2017) were selected and ten baseline cases and ten 30-day follow-up cases were then randomly selected for abstraction. The sample for the baseline and 30 day follow-up records across both years included 1,266 TAVR procedures as well as 134 procedures on mitral valve repair and mitral valve replacement. Trained nurse auditors re-abstracted preselected data elements from the medical record and these results were compared against the original registry data submitted for that procedure.

Agreement rate can be interpreted as follows based on the data assessed:

- Exceeds Expectations: agreement rate ≥ 95%
- Meets Expectation: agreement rate 85% - 95%
- Needs Improvements: agreement rate < 85%

A 95% confidence interval was calculated for each PABAK statistic to reflect sampling error and indicate a range of plausible values for the PABAK statistic. General interpretation of the PABAK statistic is similar to the KAPPA:

| **PABAK Interpretations:** | |
|---|---|
| 0.00 | Poor agreement |

| | |
|---|---|
| 0.01-0.20 | Slight agreement |
| 0.21-0.40 | Fair agreement |
| 0.41-0.60 | Moderate agreement |
| 0.61-0.80 | Substantial agreement |
| 0.81-1.00 | Almost perfect agreement |

To validate accuracy of 30-day mortality in the TVT Registry, we compared TVT Registry data linked CMS claims data from 2012-2015.

Empirical validity:

Validity of the proposed measure depends in part on the adequacy of the risk adjustment model to adjust for case mix. As such, our empirical validity testing focused on assessing consistency between the observed data and the underlying assumptions used for statistical analysis. Specifically, we created calibration plots for the overall cohort and for several pre-specified subgroups. Large discrepancies between observed and model-predicted probabilities in any of the plots would suggest that the functional form of the model was misspecified and that estimates of provider performance may be invalid. Methods and results of these analyses are provided in Section 2b4.

**2b1.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test*)

Records Eligibility Assessment (REA): REA was performed on 343 records at six sites in 2016 and 380 records at six sites in 2017 by comparing records submitted to the registry to facility-provided billing code lists. 100% of the assessed cases were valid procedures across all hospitals. In reviewing records submitted to the registry, there was one case identified that was omitted in 2016. The hospital reported it was a complex case ad was referred to clinical staff at the registry to address their questions/concerns.

Data element validity:

Agreement rates and PABAK scores are provided for variables that are categorical in the first table (Table 4) and the second table (Table 5) provides agreement rates and Pearson Correlations scores for those variables that are continuous.

**Table 4**

| Data Element | Uni-verse | Year audited | PABAK | | Agreement | | |
|---|---|---|---|---|---|---|---|
| | | | Score | 95% CI | Initial Score % | Final Score % | 10th Percentile - 90th Percentile |
| Birth Date (DOB) | 400 | 2016 | 0.98 | 0.953-1.000 | 99.0% | 99.0% | 100-100 |
| | 500 | 2017 | 0.972 | 0.94-1.000 | 96.8% | 96.8% | 97-100 |
| Sex (Sex) | 400 | 2016 | 0.97 | 0.937-1.000 | 98.5% | 98.5% | 90-100 |
| | 500 | 2017 | 0.980 | 0.96-1.00 | 99.0% | 99.0% | 98-100 |
| Permanent Pacemaker (Pacemaker) | 400 | 2016 | 0.92 | 0.867-0.973 | 96.0% | 96.3% | 90-100 |
| | 500 | 2017 | 0.964 | 0.93-1.00 | 98.2% | 98.2% | 97-100 |
| Prior PCI (PriorPCI) | 400 | 2016 | 0.91 | 0.854-0.9666 | 95.5% | 95.8% | 90-100 |
| | 500 | 2017 | 0.956 | 0.92-0.99 | 97.8% | 97.8% | 96-99 |
| Prior CABG (PriorCABG) | 400 | 2016 | 0.99 | 0.970-1.000 | 99.5% | 99.5% | 100-100 |
| | 500 | 2017 | 0.984 | 0.96-1.00 | 99.2% | 99.2% | 98-100 |
| Prior Aortic Valve Proc (PriorAorticValve) | 400 | 2016 | 0.97 | 0.929-1.000 | 98.3% | 98.3% | 90-100 |
| | 500 | 2017 | 0.992 | 0.98-1.00 | 99.6% | 99.6% | 99-100 |

| Data Element | Uni-verse | Year audited | PABAK | | Agreement | | |
|---|---|---|---|---|---|---|---|
| | | | Score | 95% CI | Initial Score % | Final Score % | 10th Percentile - 90th Percentile |
| Prior Stroke (PriorStroke) | 400 | 2016 | 0.94 | 0.893-0.987 | 97.0% | 97.3% | 90-100 |
| | 500 | 2017 | 0.928 | 0.88-0.97 | 96.4% | 96.4% | 94-99 |
| TIA (CVDTIA) | 400 | 2016 | 0.95 | 0.907-0.993 | 97.5% | 97.5% | 90-100 |
| | 500 | 2017 | 0.920 | 0.87-0.97 | 96.0% | 96.0% | 94-98 |
| Peripheral Arterial Disease (PriorPAD) | 400 | 2016 | 0.88 | 0.816-0.944 | 94.0% | 94.3% | 85-100 |
| | 500 | 2017 | 0.808 | 0.74-0.88 | 90.4% | 90.4% | 87-94 |
| Diabetes Mellitus (Diabetes) | 400 | 2016 | 0.96 | 0.914-0.996 | 97.8% | 97.8% | 90-100 |
| | 500 | 2017 | 0.928 | 0.88-0.97 | 96.4% | 96.4% | 94-99 |
| Currently on Dialysis (CurrentDialysis) | 400 | 2016 | 0.99 | 0.970-1.000 | 99.5% | 99.5% | 100-100 |
| | 500 | 2017 | 0.984 | 0.96-1.00 | 99.2% | 99.2% | 98-100 |
| Chronic Lung Disease (ChrLungD) | 400 | 2016 | 0.68 | 0.589-0.771 | 84.0% | 84.8% | 65-100 |
| | 500 | 2017 | 0.760 | 0.69-0.83 | 88.0% | 88.0% | 84-92% |
| Home Oxygen (HMO2) | 400 | 2016 | 0.92 | 0.867-0.973 | 96.0% | 96.3% | 90-100 |
| | 500 | 2017 | 0.956 | 0.92-0.99 | 97.8% | 97.8% | 96-99 |
| Hostile Chest (HostileChest) | 400 | 2016 | 0.95 | 0.900-0.990 | 97.3% | 97.5% | 90-100 |
| | 500 | 2017 | 0.920 | 0.87-0.97 | 96.0% | 96.0% | 94-98 |
| Immunocompromise (ImmSupp) | 400 | 2016 | 0.95 | 0.900-0.990 | 97.3% | 97.3% | 90-100 |
| | 500 | 2017 | 0.948 | 0.91-0.99 | 97.4% | 97.4% | 96-99 |
| Prior MI (PriorMI) | 400 | 2016 | 0.85 | 0.780-0.920 | 92.5% | 92.5% | 80-100 |
| | 500 | 2017 | 0.868 | 0.81-0.93 | 93.4% | 93.6% | 91-96 |
| Porcelain Aorta (PorcelainAorta) | 400 | 2016 | 0.96 | 0.922-0.998 | 98.0% | 98.3% | 90-100 |
| | 500 | 2017 | 0.980 | 0.96-1.00 | 99% | 99% | 98-100 |
| Atrial Fibrillation/Flutter (AFibFlutter) | 400 | 2016 | 0.93 | 0.880-0.980 | 96.5% | 96.8% | 90-100 |
| | 500 | 2017 | 0.912 | 0.86-0.96 | 95.6% | 95.6% | 93-98 |
| Five Meter Walk Test Perf (FiveMWalkTest) | 357 | 2016 | 0.82 | 0.741-0.900 | 91.0% | 91.3% | 70.7-100 |
| | 445 | 2017 | 0.717 | 0.63-0.80 | 85.8% | 86.1% | 82-90 |
| KCCQ-12 Performed (KCCQ12_Performed) | 400 | 2016 | 0.92 | 0.867-0.973 | 96.0% | 96.0% | 85-100 |
| | 500 | 2017 | 0.868 | 0.81-0.93 | 93.4% | 93.4% | 91-96 |
| MV Insufficiency (VDInsufM) | 383 | 2016 | 0.63 | 0.532-0.726 | 81.5% | 81.7% | 55-100 |
| | 500 | 2017 | 0.588 | 0.50-0.67 | 79.4% | 79.4% | 75-84 |
| Procedure Start Date (TVTProcedureStartDate) | 689 | 2016 | 1.00 | 1.000-1.000 | 100.0% | 100.0% | 100-100 |
| | 500 | 2017 | 1.00 | 1.00-1.00 | 100.0% | 100.0% | 100-100 |
| Valve Sheath Access Site (TVTAccessSite) | 357 | 2016 | 0.99 | 0.979-1.000 | 99.7% | 99.7% | 100-100 |
| | 445 | 2017 | 1.00 | 1.00-1.00 | 100.0% | 100.0% | 100-100 |
| Discharge Date (DCDate) | 400 | 2016 | 0.97 | 0.937-1.000 | 98.5% | 98.5% | 90-100 |
| | 500 | 2017 | 0.960 | 0.93-0.99 | 98.0% | 98.0% | 96-100 |
| Discharge Status (DCStatus) | 400 | 2016 | 1.00 | 1.000-1.000 | 100.0% | 100.0% | 100-100 |
| | 500 | 2017 | 0.996 | 0.98-1.00 | 99.8% | 99.8% | 99-100 |
| Follow-up Status (F_Status) | 387 | 2016 | 0.77 | 0.689-0.856 | 88.6% | 89.9% | 70-100 |
| | 302 | 2017 | 0.980 | 0.95-1.00 | 99.0% | 99.3% | 98-100 |
| Follow-up Date of Death (F_DeathDate) | 8 | 2016 | 0.50 | 0.000-1.000 | 75.0% | 75.0% | 0-100 |
| | 3 | 2017 | N/A | N/A | 33.3% | 33.3% | 0-100 |

**Table 5**

| Field Name | Universe | Year audited | Pearson Correlation | | Agreement | | |
|---|---|---|---|---|---|---|---|
| | | | Score | Lower 95% CI – Upper 95% CI | Initial Score % | Final Score % | 10th Percentile - 90th Percentile |
| Height (Height) | 400 | 2016 | 0.966 | 0.959-0.972 | 86.3% | 86.5% | 60-100 |

| Field Name | Universe | Year audited | Pearson Correlation | | Agreement | | |
|---|---|---|---|---|---|---|---|
| | | | Score | Lower 95% CI – Upper 95% CI | Initial Score % | Final Score % | 10th Percentile - 90th Percentile |
| | 500 | 2017 | 0.911 | 0.89-0.92 | 92.4% | 92.6% | 90-96 |
| Weight (Weight) | 400 | 2016 | 0.983 | 0.979-0.986 | 75.8% | 75.8% | 35-100 |
| | 500 | 2017 | 0.982 | 0.98-0.98 | 79.6% | 79.6% | 75-84 |
| Pre-Procedure Creatinine (PreProcCreat) | 400 | 2016 | 0.999 | 0.999-0.999 | 92.3% | 92.3% | 80-100 |
| | 500 | 2017 | 0.986 | 0.98-0.99 | 85.0% | 85.0% | 85-92 |
| Left Ventricle Ejection Fraction (LVEF) | 398 | 2016 | 0.963 | 0.956-0.970 | 77.9% | 78.6% | 40-100 |
| | 500 | 2017 | 0.931 | 0.92-0.94 | 68.2% | 68.2% | 63-74 |

The incidence of death captured post discharge up to 30 days is infrequent, making it difficult to validate in audits. To validate accuracy of 30-day mortality in the TVT Registry, we compared TVT Registry data linked CMS claims data from 2012-2015 (refer to the yellow highlighting in Table 6 below).

**Table 6**

| Variable | Level | Overall (N=41582) | | 2012 (N=4656) | | 2013 (N=9104) | | 2014 (N=16389) | |
|---|---|---|---|---|---|---|---|---|---|
| **Using Registry Only Data** | | | | | | | | | |
| 30 Day Death (Among non-missing) | No | 34884 | 93.72 | 3901 | 92.53 | 7734 | 92.92 | 14029 | 93. |
| | Yes | 2337 | 6.28 | 315 | 7.47 | 589 | 7.08 | 909 | 6. |
| 30 Day Death (Among entire registry) | Missing | 4361 | 10.49 | 440 | 9.45 | 781 | 8.58 | 1451 | 8. |
| | No | 34884 | 83.89 | 3901 | 83.78 | 7734 | 84.95 | 14029 | 85. |
| | Yes | 2337 | 5.62 | 315 | 6.77 | 589 | 6.47 | 909 | 5. |
| **Using CMS Only Data** | | | | | | | | | |
| 30 Day Death (Among linked procedures) | No | 27607 | 94.03 | 3092 | 92.63 | 6076 | 93.23 | 10867 | 94. |
| | Yes | 1752 | 5.97 | 246 | 7.37 | 441 | 6.77 | 661 | 5. |
| **Using CMS & Registry Data** | | | | | | | | | |
| 30 Day Death Discrepancy (Among linked procedures) | No | 29222 | 99.53 | 3320 | 99.46 | 6486 | 99.52 | 11476 | 99. |
| | Yes | 137 | 0.47 | 18 | 0.54 | 31 | 0.48 | 52 | 0. |
| 30 Day Death Discrepancy: Reg Y, CMS N (Among linked procedures) | No | 29334 | 99.91 | 3337 | 99.97 | 6511 | 99.91 | 11516 | 99. |
| | Yes | 25 | 0.09 | 1 | 0.03 | 6 | 0.09 | 12 | 0. |
| 30 Day Death Discrepancy: Reg N, CMS Y (Among linked procedures) | No | 29247 | 99.62 | 3321 | 99.49 | 6492 | 99.62 | 11488 | 99. |
| | Yes | 112 | 0.38 | 17 | 0.51 | 25 | 0.38 | 40 | 0. |

Empirical validity:

For results of empirical validity testing of the case mix adjustment model see Section 2b4. Additional empirical validity testing was unnecessary because the importance of a low mortality rate is undisputed and it can be measured directly.

**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?*)

Record eligibility assessment: This assessment confirms a high degree of accuracy in reporting eligible cases at each hospital. There was no evidence of purposeful under reporting of eligible cases.

Data element validity:

All of the data elements had agreement rates that met or exceeded expectations and PABAK scores of moderate to excellent agreement over multiple years. While we are not able to assess all model variables because of competing regulatory requirements for post approval studies in the TVT Registry, 24 of the 41 model variables in addition to the 6 critical data elements for the numerator and denominator are provided. We are re-evaluating adding additional risk model variables in future years.

The agreement rates of Follow-up Status were lower than expected. Results were left skewed with one site having 0.0% agreement. Thirty-four of the 44 mismatches were a result of no 30-day follow-up status being submitted to the registry despite documentation that supported a 30-day follow-up status of alive or deceased status. The other ten had submitted a follow-up status of alive or dead and there was no documentation present for the auditor to validate the answer during review.

The agreement rates for Follow-up Date of Death varied from 0.00% to 100.00% with a standard deviation of 47.4%. Due to the limited number of sites with data, variation, skewing, or outliers could not be determined. One of the mismatches was due to lack of supporting documentation for the date the death occurred. It is noted in our data quality reports that date of death is not documented in 0.9% of follow-up deaths between January 1 2015 and December 31, 2017. In these cases, the follow-up assessment date is used as a surrogate to determine date of death.

In addition, the Five Meter Walk Test Performed had almost perfect agreement and PABAK results, the results from the 2016 audit were lower than the other data elements. ACC and STS continue to work with hospitals to encourage them to increase the number of patients for whom follow-up data is collected as well as perform and document the Five Meter Walk test. It was also noted that the disagreement was due to either (1) the test being performed in a timeframe longer than the target value (30 days pre-procedure), or (2) the test being documented in feet, not meters. To address these, staff have widened the acceptable target value for this test in an upcoming dataset upgrade and reinforce the assessment should be documented in meters (not feet).

Two of the conditions (MV insufficiency and chronic lung disease) included as variables in the risk model have lower agreement rates and PABAK in 2016 but were improved in the 2017 audit. Physician members have commented that the agreement rates for valve insufficiency also varies in core lab validations of patients in clinical trials. In addition, for transcatheter procedures (compared to surgical procedures) the severity of chronic lung disease is not well documented. ACC and STS continue to refine definitions and work with hospitals to improve documentation practices.

Because the incidence of death captured post discharge up to 30 days is infrequent, it is difficult to validate in audits. A separate comparison of the TVT Registry data linked to CMS claims data from 2012-2015 demonstrated that across 3.5 years, 99.6% of the 29,247 patient records had no discrepancy.

Empirical validity:

As discussed in Section 2b4, results suggest that the risk adjustment model is well calibrated and suitable to

adjust for case mix.

**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*)

As noted in Section 1.6, the analysis was restricted to hospitals with ≥90% complete non-missing data for 30-day mortality status, baseline KCCQ-12 score, and baseline gait speed. To explore how included and excluded TVT centers might differ, we compared site-level factors, patient-level characteristics, and outcomes between included and excluded sites using 2-sided Wilcoxon rank-sum tests for median values and standardized differences (a >10% difference is considered clinically relevant) for categorical variables. In addition, we discuss the clinical importance of requiring completion of the KCCQ and gait speed and why the threshold was set at >=90%. We also provide additional information on the differences between sites that were excluded vs. included and how the number of sites who meet the minimum data completeness requirements has improved over time.

**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*)

As shown in the Table 7, there were no significant differences in teaching status, bed size, or annual TAVR procedural volume between included and excluded sites. There were few meaningful differences between patients from included and excluded sites, with patients from included sites being less likely to be of nonwhite race or Hispanic ethnicity (6.4% vs. 11.1%, standardized difference 17%) and more likely to have a tricuspid aortic valve (94.7% vs. 87.6%, standardized difference 25%). The rate of death at 30 days was also similar between included and excluded sites (4.7%vs. 5.1%, standardized difference 2%).

**Table 7. Characteristics of patients treated at included versus excluded sites**

| | **Included Sites site n=188 patient n=22,506** | **Excluded Sites site n=262 patient n=38,264** | **Standardized Difference or p-value[1]** |
|---|---|---|---|
| Teaching hospital | 56.4% | 62.6% | 0.185 |
| Number of beds (median [IQR]) | 452 (355-626) | 516 (359-685) | 0.172 |
| Annual TAVR volume (median [IQR]) | 43 (32-63) | 48 (33-77) | 0.057 |
| Death within 30 days | 4.7% | 5.1% | 2% |
| Length of stay (days; median [IQR]) | 5 (3-8) | 5 (3-8) | 0.890 |
| Age | | | 5% |
| <75 years | 20.8% | 22.4% | |
| 75-84 years | 36.8% | 37.0% | |
| ≥85 years | 42.4% | 40.6% | |
| Female sex | 48.4% | 47.3% | 2% |
| Non-white race or Hispanic ethnicity | 6.4% | 11.1% | 17% |
| Body Surface Area | | | 1% |
| <1.80 $m^2$ | 41.5% | 42.1% | |

| | | | |
|---|---|---|---|
| 1.80-2.19 $m^2$ | 48.9% | 48.7% | |
| ≥2.20 $m^2$ | 9.6% | 9.3% | |
| Prior myocardial infarction | 25.1% | 24.8% | 1% |
| Prior coronary stenting | 34.8% | 35.0% | 1% |
| Prior coronary bypass surgery | 27.6% | 27.3% | 1% |
| Prior aortic procedure | 12.8% | 14.6% | 5% |
| Prior non-aortic procedure | 2.3% | 2.7% | 2% |
| Left main stenosis ≥50% | 10.0% | 10.3% | 1% |
| Proximal LAD stenosis ≥70% | 19.8% | 19.5% | 1% |
| Prior stroke or transient ischemic attack | 19.2% | 18.4% | 2% |
| Carotid stenosis | 23.4% | 19.7% | 9% |
| Peripheral artery disease | 30.8% | 30.2% | 1% |
| Atrial fibrillation or flutter | 41.3% | 40.3% | 2% |
| Conduction defect | 36.4% | 34.9% | 3% |
| Implantable defibrillator | 3.9% | 4.7% | 4% |
| Diabetes mellitus | 37.5% | 37.5% | 0% |
| Severe chronic lung disease | 14.0% | 13.3% | 2% |
| Home oxygen use | 12.3% | 11.2% | 4% |
| Current smoker | 5.8% | 5.6% | 1% |
| Renal Function | | | 1% |
| GFR ≥60 mL/min/1.73$m^2$, no dialysis | 52.9% | 52.7% | |
| GFR 30-59 mL/min/1.73$m^2$, no dialysis | 41.9% | 41.7% | |
| GFR <30 mL/min/1.73$m^2$, no dialysis | 5.2% | 5.6% | |
| Current dialysis | 3.7% | 4.4% | 4% |
| Ejection Fraction | | | 4% |
| <35% | 11.1% | 12.3% | |
| 35-54% | 24.7% | 23.7% | |
| ≥55% | 64.2% | 63.9% | |
| Hemoglobin <10 g/dL | 14.5% | 16.7% | 6% |
| Platelet <100,000 /mL | 4.0% | 4.7% | 3% |
| Hostile chest | 7.0% | 8.2% | 5% |
| Porcelain aorta | 5.8% | 5.4% | 2% |
| Endocarditis | 0.8% | 1.1% | 3% |
| Degenerative aortic valve | 95.6% | 94.3% | 6% |
| Tricuspid aortic valve | 94.7% | 87.6% | 25% |
| Aortic insufficiency ≥ moderate | 20.3% | 20.4% | 0% |
| Mitral insufficiency ≥ moderate | 29.7% | 28.2% | 3% |
| Tricuspid insufficiency ≥ moderate | 24.4% | 23.8% | 1% |
| Non-femoral access | 21.3% | 18.9% | 6% |
| Acuity Category | | | 4% |
| (1) Elective | 89.0% | 87.8% | |

| | | | |
|---|---|---|---|
| (2) Urgent | 7.7% | 8.4% | |
| (3) Pre-procedure shock | 2.9% | 3.3% | 1% |
| (4) Emergent/salvage | 0.4% | 0.5% | 1% |
| 5-Meter Walk Test | | | |
| Missing % | 3.3% | 26.1% | |
| Unable to walk | 11.9% | 13.3% | 4% |
| Walk speed by quartiles | | | 12% |
| Q1 (speed < 0.417 m/s) | 26.4% | 31.3% | |
| Q2 (speed 0.417-0.624 m/s) | 29.4% | 27.0% | |
| Q3 (speed 0.625-0.788 m/s) | 17.4% | 15.0% | |
| Q4 (speed ≥ 0.789 m/s) | 26.8% | 26.6% | |
| KCCQ | | | |
| Missing % | 2.2% | 17.2% | |
| KCCQ by quartiles | | | 5% |
| Q1 (KCCQ <23.96) | 25.7% | 27.6% | |
| Q2 (KCCQ 23.96-40.39) | 24.5% | 24.0% | |
| Q3 (KCCQ 40.10-58.32) | 23.5% | 22.5% | |
| Q4 (KCCQ ≥58.33) | 26.4% | 25.8% | |

[1] Standardized differences shown for patient-level variables due to large sample size. Greater than 10% is generally considered a meaningful difference between groups. P-values for site-level characteristics and for patient length of stay were derived from Wilcoxon Rank Sum Test are shown for site-level characteristics.

Additional comments on the model inclusion/exclusion criteria requiring documentation of KCCQ and gait speed:

Clinical importance of KCCQ and gait speed: Physician leaders and model developers feel it is important to use assessment of health status (via KCCQ-12) and frailty (via 5-meter walk test) in our risk models (especially for this patient population). Documentation of baseline KCCQ is required to meet the CMS "Coverage with Evidence Determination" for TAVR, describing and monitoring symptoms, functional status and quality of life for patients with heart failure. Worse baseline KCCQ scores are associated with higher risk for mortality after TAVR. In addition, slower gait speed, which is an important marker of frailty, independently predicts risk of mortality after TAVR.

Determination of >=90% completeness threshold: In 2016, model developers reviewed different data completeness threshold's impact on the number of sites and patients included (see table 8). Based on a review of data completeness at different thresholds, they felt that we should limit analysis and hospital feedback to sites with >=90% completeness on these variables to improve internal validity. KCCQ and gait speed were imputed to the median for patient records that had missing data. Imputation slightly penalizes sites because they do not benefit from full risk adjustment (patients with missing data may appear to be less sick than they actually are). Since 90% is the standard data quality completeness threshold for all data elements in risk models, we felt this bar of 90% was reasonable, given the expectations to perform these assessments.

Differences between sites that were included/excluded: As shown in the Table 7, there were no significant differences in teaching status, bed size, or annual TAVR procedural volume between included and excluded sites. There were few meaningful differences between patients from included and excluded sites, with patients from included sites being less likely to be of nonwhite race or Hispanic ethnicity (6.4% vs. 11.1%, standardized difference 17%) and more likely to have a tricuspid aortic valve (94.7% vs. 87.6%, standardized difference 25%).

The rate of death at 30 days was also similar between included and excluded sites (4.7%vs. 5.1%, standardized difference 2%).

Improvement: We have expected a slow improvement of the # of sites included over time from initial development (since the model reports a “rolling 3 year” timeframe, it takes a while for a site to catch up on data completeness). There has been an improvement in the # of sites included (188 hospitals in initial development; 301 sites in the 2018q4 published outcome reports). We continue to monitor this in the future.

**Table 8**

**Completeness: TVTR 30 Day Mortality/Some In-hospital KCCQ-12 Component/Some 5 Meter Walk Time**

| Percent of Records with Complete Data | | | | | |
|---|---|---|---|---|---|
| Site Completeness Threshold | # Sites Included | # Records Included | Complete for TVTR 30 Day Death[1] | Complete for Some In-hospital KCCQ-12[2] | Complete for Some 5m Walk Time[3] |
| ≥ 0% | 450 | 60770 | 90.0% | 88.3% | 82.3% |
| ≥ 10% | 435 | 59781 | 90.0% | 89.4% | 83.4% |
| ≥ 20% | 429 | 58828 | 90.1% | 90.0% | 84.5% |
| ≥ 30% | 426 | 58362 | 90.6% | 90.1% | 84.5% |
| ≥ 40% | 416 | 57452 | 90.9% | 90.2% | 85.0% |
| ≥ 50% | 397 | 54175 | 91.8% | 91.5% | 87.1% |
| ≥ 60% | 375 | 50943 | 92.3% | 92.7% | 88.7% |
| ≥ 70% | 345 | 45472 | 93.0% | 93.7% | 91.1% |
| ≥ 80% | 281 | 34747 | 94.4% | 96.1% | 94.0% |
| ≥ 90% | 188 | 22506 | 96.2% | 97.8% | 96.7% |
| ≥ 100% | 19 | 402 | 100.0% | 100.0% | 100.0% |

[1]TVTR 30 Day Death is non-missing (see specs for definition)
[2]Some In-hospital KCCQ-12 (#5170-5181) component value is non-missing. There are multiple questions incorporated into the KCCQ-12
measurement. This table reports subjects in which atleast some of these questions are answered.
[3]Some 5m Walk Time (#5090, #5095, #5100) atleast one entry is non-
missing or patient is Unable to Walk (#5085 = 2)

**2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results?** (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The goals of the site-level exclusion are: (1) to improve internal validity by minimizing the amount of missing data that must be imputed or excluded and (2) to incentive TVT participants to obtain complete data. We acknowledge that the interpretation of the measure results is impacted by this exclusion. In particular, a hospital's odds ratio must be interpreted as a comparison between the hospital of interest and the set of TVT centers that were included in the analysis. Although such a comparison is less desirable than a comparison between a hospital and all US TAVR centers, it is still internally valid as a comparison with the centers included. In addition, our analysis did not detect large or important differences in characteristics between the centers that were included and excluded. Importantly, the model will be re-estimated for the TVT registry participant feedback report each quarter. In addition, we noted a higher proportion of TAVR centers included in the TVT outcomes report (see table 10) as the proportion of sites with complete data improves.

**Table 9 – Comparison of hospitals INCLUDED in different timeframes of TESTING:**

| Initial Development | Current Testing Cohort |
|---|---|
| **June 2013 – May 2016** | **April 2015 – March 2018** |
| 188 hospitals | 264 hospitals |

**Table 10 – Comparison of hospitals included in different timeframes of PUBLISHED OUTCOMES REPORTS:**

| Initial timeframe published in site outcome report | Most recent timeframe published in site outcome report |
|---|---|
| **Jan 2015-Dec 2017** | **Jan 2016- Dec 2018** |
| 253 hospitals | 301 hospitals |

**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** 41 **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.**

Case mix adjustment was implemented using a hierarchical logistic regression mode with site-specific random intercept parameters. Covariates in the model are listed in Section 2b4.4a below. For additional details, please see attached article:

Arnold SV, O'Brien SM, Vemulapalli S, Cohen DJ, Stebbins A, Brennan JM, Shahian DM, Grover FL, Holmes DR, Thourani VH, Peterson ED, Edwards FH; STS/ACC TVT Registry. Inclusion of Functional Status Measures in the Risk Adjustment of 30-Day Mortality After Transcatheter Aortic Valve Replacement: A Report From the Society of Thoracic Surgeons/American College of Cardiology TVT Registry. JACC Cardiovasc Interv. 2018 Mar 26;11(6):581-589. doi: 10.1016/j.jcin.2018.01.242.

**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

**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?

Covariates for the models included all factors from the original TAVR in-hospital mortality model, which were selected on the basis of clinical judgement (12). In addition, baseline KCCQ-12 scores and gait speed were included. Because of conceptual overlap with these measures and to minimize potential for gaming, New York Heart Association (NYHA) functional class IV was removed as a covariate.

Because the purpose of these models is for risk adjustment of outcomes for site reporting, all covariates deemed clinically relevant were retained in these nonparsimonious models. Linearity for all continuous

variables was tested using restricted cubic splines, and variables with nonlinear relationships with the outcome were categorized, as appropriate.

Our framework for selecting covariates was based on the statistical literature on treatment effect estimation in non-randomized observational studies. In our context, the treatment effect of interest is the effect of undergoing TAVR at a particular hospital compared to undergoing TAVR at a hypothetical average TAVR hospital. Valid estimation of this treatment effect requires the assumption that outcome differences are unconfounded conditional on a set of pre-TAVR baseline covariates. This assumption means that, within blocks of patients having identical values of pre-TAVR covariates, patients at each hospital are like a random sample from a common patient population. Although the unconfoundedness assumption is unlikely to be literally tre in a non-randomized observational study, the risk of encountering large violations of the assumption can be minimized by adjusting for a large number of pre-TAVR covariates. Thus, our modeling strategy was non-parsimonious and did not select or remove covariates on the basis of their statistical significance in a model predicting outcomes.

Refer to section 1.8 for a detailed description of social risk factors considered for this model.

**2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:**
**X Published literature**
**x Internal data analysis**
☐ **Other (please describe)**

**2b3.4a. What were the statistical results of the analyses used to select risk factors?**

List of covariates in the final model were:

| | | |
|---|---|---|
| Age | Prior peripheral artery disease | # prior cardiac operations |
| BSA | Current/recent smoker | Prior aortic procedure |
| Sex | Diabetes | Prior other valve procedure |
| Race/ethnicity | | Aortic etiology |
| eGFR | Atrial fibrillation/flutter | Valve morphology |
| Dialysis | Conduction defect | Aortic insufficiency |
| Ejection fraction | Chronic lung disease | Mitral insufficiency |
| Hemoglobin | Home oxygen | Tricuspid insufficiency |
| Platelet count | Hostile chest | Acuity status |
| Procedure date | Porcelain aorta | Cardiogenic shock |
| LMD ≥ 50% | Access site | Cardiac arrest w/in 24 hours |
| Proximal LAD ≥ 70% | Pacemaker | Pre-procedure inotropes |
| Prior MI | Previous ICD | Mechanical assist device |
| Endocarditis | Prior PCI | Carotid stenosis |
| Gait speed | Prior CABG | Prior TIA/stroke |
| Baseline KCCQ-12 | | |

Covariates were selected a prior and were not removed on the basis of their statistical significance.

**Table 11. TAVR 30-day mortality risk adjustment model Covariates, Odds Ratio and P Values**

| Covariate | Odds Ratio (95% CI) | P Value |
|---|---|---|
| Age (per 5 years when ≤75) | 0.91 (0.82-1.01) | 0.089 |
| Age (per 5 years when >75) | 1.19 (1.11-1.28) | <0.001 |

| | | |
|---|---|---|
| Sex (female at BSA 1.7 $m^2$ vs male at BSA 1.9 $m^2$) | 1.00 (0.87-1.16) | 0.959 |
| Race (non-white race or Hispanic ethnicity) | 1.15 (0.87-1.52) | 0.341 |
| Body surface area (per 1 $m^2$ for male) | 0.33 (0.21-0.54) | <0.001 |
| Body surface area (per 1 $m^2$ for female) | 0.45 (0.28-0.71) | <0.001 |
| Prior myocardial infarction | 1.21 (1.04-1.42) | 0.015 |
| Prior coronary stenting | 0.89 (0.77-1.03) | 0.117 |
| Prior coronary bypass surgery | 0.76 (0.57-1.03) | 0.080 |
| Prior cardiac operations, (1 vs. 0) | 0.99 (0.75-1.31) | 0.953 |
| Prior cardiac operations, (2+ vs. 0) | 0.80 (0.51-1.27) | 0.349 |
| Prior aortic valve procedure | 1.11 (0.92-1.33) | 0.272 |
| Prior non-aortic valve procedure | 0.69 (0.42-1.15) | 0.156 |
| Left main stenosis ≥50% | 1.34 (1.07-1.67) | 0.011 |
| Proximal LAD stenosis ≥70% | 1.08 (0.90-1.31) | 0.409 |
| Prior stroke or transient ischemic attack | 0.91 (0.77-1.07) | 0.250 |
| Carotid stenosis | 1.09 (0.94-1.27) | 0.272 |
| Peripheral artery disease | 1.23 (1.06-1.41) | 0.006 |
| Atrial fibrillation or flutter | 1.13 (0.99-1.29) | 0.081 |
| Conduction defect | 0.97 (0.85-1.12) | 0.703 |
| Pacemaker | 0.92 (0.77-1.10) | 0.345 |
| Implantable defibrillator | 1.18 (0.85-1.65) | 0.316 |
| Diabetes mellitus | 0.89 (0.77-1.02) | 0.100 |
| Severe chronic lung disease | 1.15 (0.96-1.38) | 0.143 |
| Home oxygen | 1.54 (1.28-1.84) | <0.001 |
| Current smoker | 0.92 (0.70-1.23) | 0.586 |
| GFR (per 5 mL/min/1.73m2) | 0.96 (0.94-0.97) | <0.001 |
| Current dialysis vs no dialysis and GFR=90 | 2.04 (1.49-2.79) | <0.001 |
| Ejection fraction (per 5%) | 0.99 (0.96-1.01) | 0.306 |
| Hemoglobin (per 1 g/dL) | 0.98 (0.94-1.02) | 0.353 |
| Platelet count (per 10,000 when ≤200,000) | 0.97 (0.95-0.99) | 0.007 |
| Platelet count (per 10,000 when >200,000) | 1.02 (1.00-1.03) | 0.014 |
| Hostile chest | 1.25 (0.97-1.61) | 0.088 |
| Porcelain aorta | 1.14 (0.88-1.47) | 0.317 |
| Endocarditis | 0.63 (0.27-1.51) | 0.303 |
| Aortic etiology (degenerative vs other) | 0.89 (0.66-1.21) | 0.467 |
| Valve morphology (tricuspid vs other) | 1.12 (0.82-1.51) | 0.486 |
| Aortic insufficiency (moderate/severe) | 0.86 (0.73-1.02) | 0.080 |
| Mitral insufficiency (moderate/severe) | 0.92 (0.79-1.06) | 0.242 |
| Tricuspid insufficiency (moderate/severe) | 1.49 (1.29-1.73) | <0.001 |
| Non-femoral access | 1.89 (1.61-2.21) | <0.001 |
| Acuity category 2 | 1.67 (1.37-2.04) | <0.001 |
| Acuity category 3 | 1.89 (1.42-2.52) | <0.001 |
| Acuity category 4 | 5.12 (2.94-8.93) | <0.001 |
| Unable to walk vs able to walk and speed = 1$^{st}$ % | 1.27 (1.02-1.58) | 0.036 |
| Gait speed (per 0.2 m/sec) | 0.95 (0.89-1.02) | 0.146 |
| Baseline KCCQ score (per 25 points) | 0.82 (0.76-0.89) | <0.001 |
| Date of procedure (per 30 day) | 0.99 (0.98-0.99) | <0.001 |

GFR, glomerular filtration rate; LAD, left anterior descending; KCCQ, Kansas City Cardiomyopathy Questionnaire

In order to explore disparities, we modified the measure's hierarchical model to include indicator variables for

black race, other non-white race, Hispanic ethnicity, and participation in Medicaid. We performed this analysis using data from June 2013 to May 2016 (188 hospitals) and using data from April 2015 to March 2018 (264 hospitals). In order to accommodate these variables, we removed an existing related variable that was defined as "non-white race or Hispanic ethnicity". Results are summarized by in the form of odds ratios in Table 12. For each variable in each time period, the 95% confidence interval around the odds ratio overlaps with the null value of 1.0. This implies that there was no statistically significant association between these variables and 30-day mortality after adjusting for other factors in the hierarchical model (p>0.05 for each variable below).

**Table 12**

| | **Estimated Odds Ratios (95% Confidence Intervals)** | |
|---|---|---|
| **Variable** | **June 2013 – May 2016** | **April 2015 – March 2018** |
| Medicaid | 0.93 (0.69 - 1.24) | 1.05 (0.83 - 1.32) |
| Black race (versus white) | 0.73 (0.47 - 1.14) | 0.91 (0.67 - 1.23) |
| Other non-white race (versus white) | 0.68 (0.36 - 1.30) | 0.85 (0.54 - 1.34) |
| Hispanic ethnicity | 1.23 (0.82 - 1.84) | 0.82 (0.59 - 1.16) |

**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.**

As noted in section 1.8, we included all covariates (except for one) from the existing TAVR in-hospital mortality model and did not exclude covariates based on their apparent statistical significance. Our goal was to adjust for all potentially relevant pre-treatment confounder variables and for this purpose it was not critical to distinguish whether a variable was directly or indirectly measuring a patient's SDS. If the prevalence of a pre-treatment prognostic factor varies across hospitals then it is a potential confounder. In the development of the prior TAVR in-hospital mortality we noted wide between-hospital variation in the prevalence of several variables that are known to be associated with a patient's SDS. These include the frequency of non-White race or Hispanic ethnicity (range across hospitals 0% to 70%), female sex (range across hospitals 34% to 63%), and age ≥65 years (range across hospitals 23% to 68%). Adjusting for these variables was regarded as desirable for face validity and to reduce confounding.

**2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model <u>or</u> stratification approach** (*describe the steps—do not just name a method; what statistical analysis was used*)

Case mix adjustment was implemented using a hierarchical logistic regression mode with site-specific random intercept parameters. Because the purpose of the model was to adjust for confounding, we preferred a non-parsimonious model and retained all covariates that were deemed to be potential confounders. Linearity for all continuous covariates was tested using restricted cubic splines, and covariates with nonlinear relationships with the outcome were categorized or modeled with spline terms. Model calibration and discrimination were assessed using a split-sample technique (70/30 split) by calculating the C-statistic and comparing observed versus expected mortality rates across deciles of predicted risk overall and within pre-specified subgroups.

*Provide the statistical results from testing the approach to controlling for differences in patient*

*characteristics (case mix) below*.
***If stratified, skip to 2b3.9***

**2b3.6. Statistical Risk Model Discrimination Statistics** (*e.g., c-statistic, R-squared*)**:**

The C statistic in the validation sample was 0.703.

**2b3.7. Statistical Risk Model Calibration Statistics** (*e.g., Hosmer-Lemeshow statistic*):

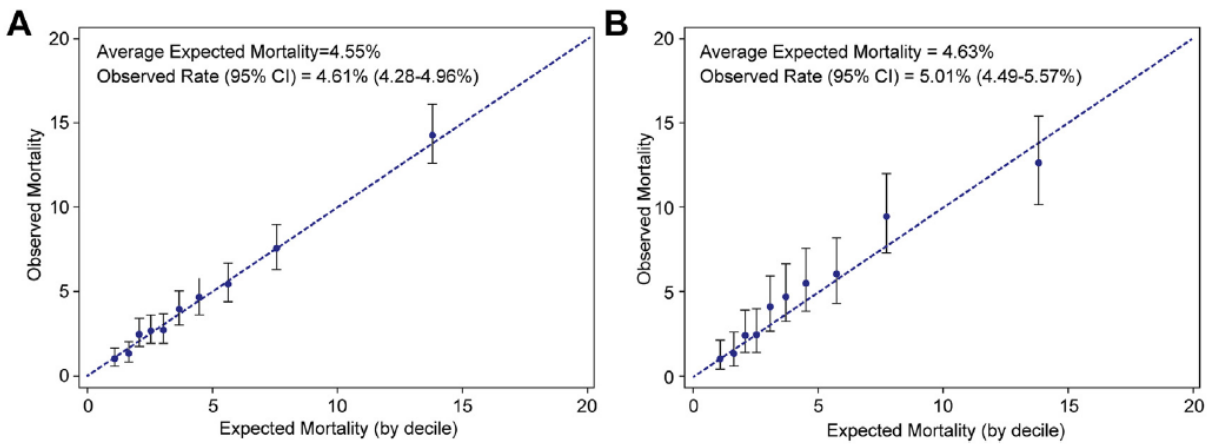
N/A. Calibration was assessed graphically.

**2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves**:

Figure 1 below displays calibration plots from the development and validation samples. Figures illustrating calibration in pre-specified subgroups (age, sex, ejection fraction, NYHA class, prior aortic procedure) are available at the end of this testing document (Supplemental Figure 1) .

**Figure 1**

**FIGURE 1 Calibration of the 30-Day Risk Adjustment Model in the Derivation (A) and Validation (B) Cohorts**




Deciles of predicted odds are plotted against the observed odds of death at 30 days after transcatheter aortic valve replacement, with 95% confidence intervals (CIs). The **dashed line** represents perfect calibration.

**2b3.9. Results of Risk Stratification Analysis**:

N/A

**2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)?** (i.*e., what do the results mean and what are the norms for the test conducted*)

Calibration of the observed and expected data was excellent in the development sample and remained good in the validation sample. Calibration was also good in several pre-defined subgroups (see supplemental figure 1, page 24-25 for results based on age, sex, ejection fraction, NYHA and prior aortic valve replacement.

**2b3.11. Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

**2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE**

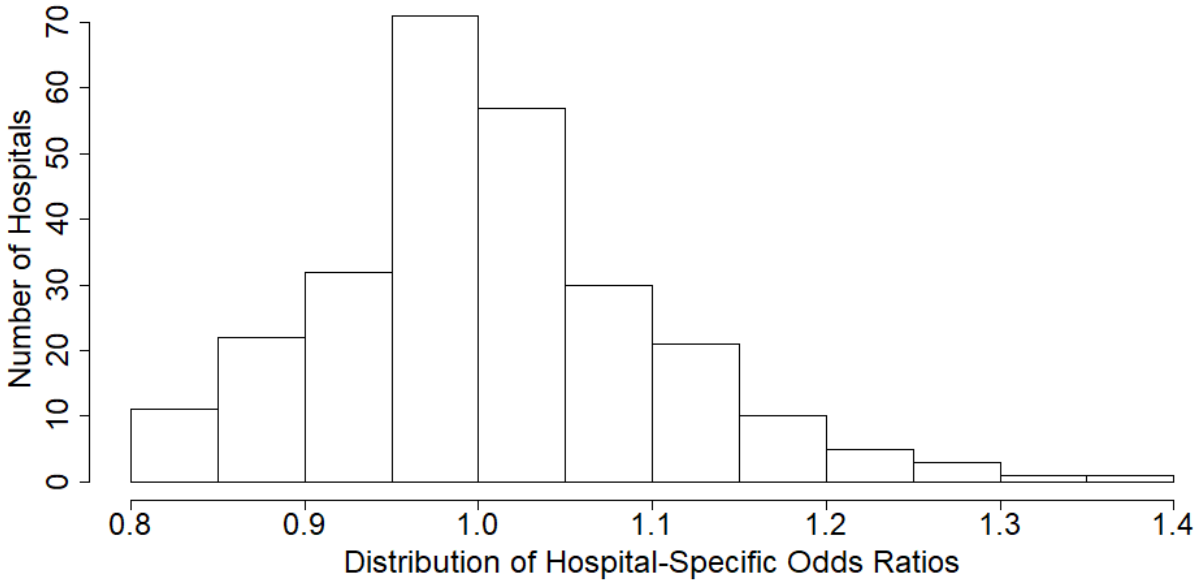
**2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified** (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Analyses for this section were based on data from April 2015 – March 2018. Using the model's estimated variance parameter, we calculated the odds ratio comparing a patient's predicted odds of dying if treated by a hospital 1 standard deviation above the mean relative to a hospital 1 standard deviation below the mean. We also created a histogram displaying the distribution of hospital-specific risk-adjusted mortality results (odds ratios).

**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*)

According to the model, a patient's predicted odds of dying was 50% higher if treated by a hospital 1 standard deviation above the mean compared with a hospital 1 standard deviation below the mean (odds ratio =1.5). Hospital-specific estimated odds ratios ranged from 0.81 to 1.40. Below is a histogram of these hospital-specific odds ratio point estimates.

Table 13




**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?*)

Analysis of data from April 2015 to March 2018 suggest that there is a substantial amount of true signal variation. The ability to detect high versus low performers is partly a function of hospital-specific sample sizes, and these may increase in the future as the number of patients undergoing TAVR increases.

In assessing the incremental utility of the 5 meter walk test[1] and baseline KCCQ score[3], model developers performed a comparison including gait speed and KCCQ versus physician reported NYHA functional class IV) to assess the improvement in risk adjustment with inclusion of these variables. A comparison of the two groups were performed using Akaike information criterion (AIC), where smaller AIC values indicate a better fit of the model. The model that included NYHA instead of gait speed had an AIC of 7,718.84 versus 7,695.64 in the KCCQ/gait speed model. The model that included NYHA instead of gate speed and KCCQ had a C statistic of 0.708 (vs 0.713). Both assessments led to the conclusion that KCCQ/gait speed model better fit the data[2].

[1]Alfredsson J, Stebbins A, Brennan JM, et al. Gait speed predicts 30-day mortality after transcatheter aortic valve replacement: results from the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. Circulation 2016;133:1351–9.

[2]Arnold, S.V. et al. Measures in the Risk Adjustment of 30-Day Mortality After Transcatheter Aortic Valve Replacement: A Report From the STS/ACC TVT Registry JACC: Cardiovascular Interventions Volume 11, Issue 6, 26 March 2018, Pages 581-589

[3]Arnold SV, Spertus JA, Vemulapalli S, et al. Association of patient-reported health status with long-term mortality after transcatheter aortic valve replacement: report from the STS/ACC TVT Registry. Circ Cardiovasc Interv 2015;8.

---

**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*)

Variables used in the analysis were highly complete with the exception of 30-day mortality status, KCCQ-12 score, and gait speed. Missing values for these and other variables were imputed to the most common category of categorical variables and to the median or subgroup-specific median of continuous variables. To reduce bias from missing data, only sites with ≥90% complete data for 30-day mortality, KCCQ-12 score, and gait speed were eligible to be included and receive a risk-adjusted 30-day mortality estimate. As a result of this exclusion, a hospital's estimated odds ratio must be interpreted as a comparison between the hospital's results and the set of TVT centers that were included in the analysis. To explore how included and excluded centers might differ, we compared site-level factors, patient-level characteristics, and outcomes between included and excluded sites using 2-sided Wilcoxon rank-sum tests for median values and standardized differences (a >10% difference is considered clinically relevant) for categorical variables. Results of this analysis are reported in Section 2b3.2 above.

**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*)

Table 14 below shows the number of sites and patients remaining after excluding sites with >10% missing data on 30-day mortality status, KCCQ-12, or gait speed during April 2015 – March 2018. Based on these data, the exclusion of sites with >10% missing data excluded 298 of 562 (=53%) of potentially eligible sites and 70,613 of 121,874 (=58%) of potentially eligible patients in the recent data.

**Table 14**

| Population Restriction | Patients Remaining | Sites Remaining |
|---|---|---|
| All TAVR during 3-year time frame | 122176 | 562 |
| Eligible sites | 51261 | 264 |
| Non-missing 30-day mortality status | 49182 | 264 |

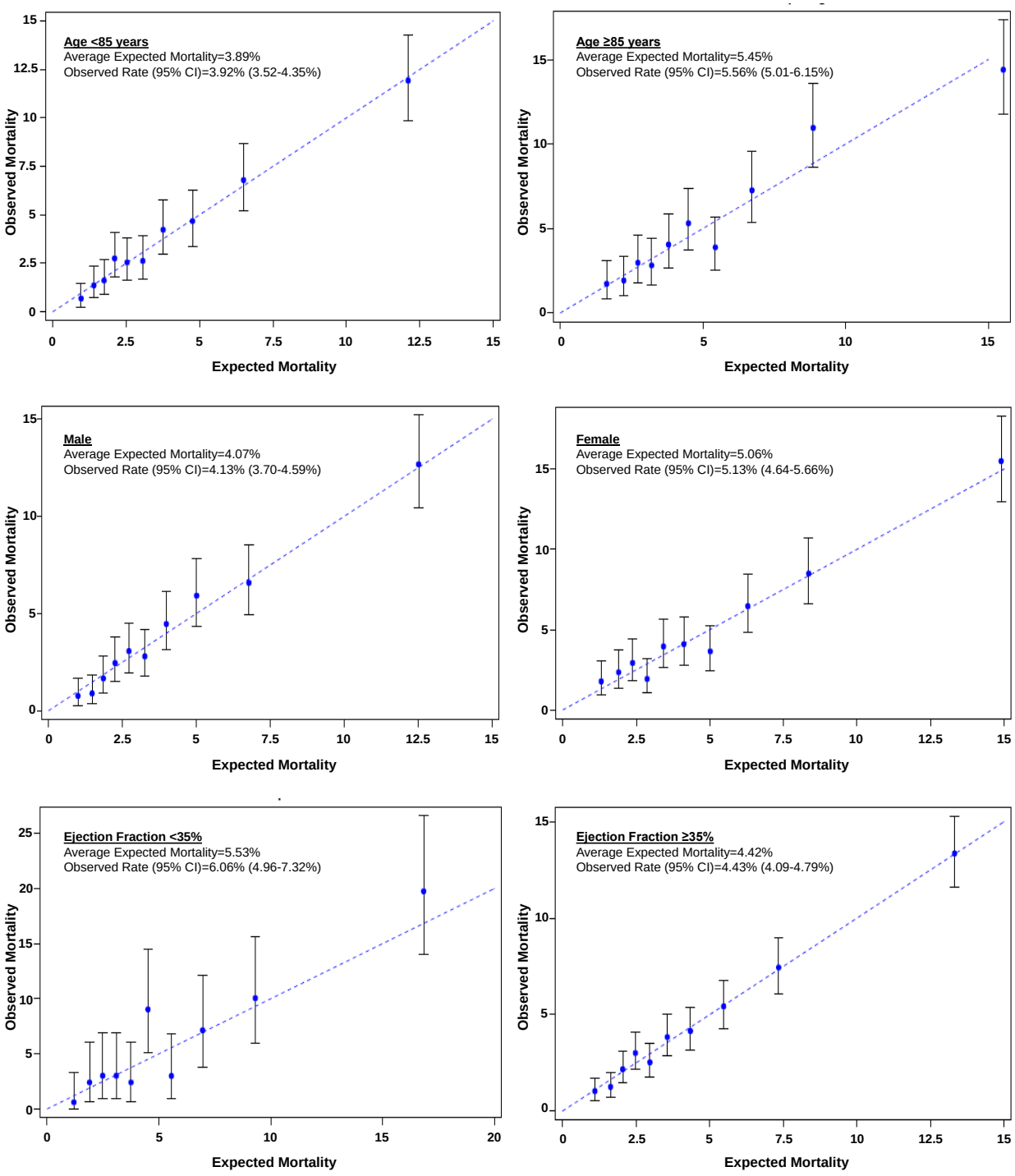
After excluding sites with high rates of missing data, the percent of cases with non-missing 30-day mortality status was 96% (49182/51261). In the final analysis cohort, missing data was 11.7% for baseline KCCQ-12 scores and 17.7% for baseline gait speed. As patients with missing data for KCCQ-12 and gait speed tend to be slightly sicker than those with collected data (with lower KCCQ-12 scores and slower gait speeds, on average), imputing missing KCCQ-12 and gait speed to the median essentially inserted a slight negative bias by making the missing patients appear less sick than they actual are. This will essentially slightly penalize sites with

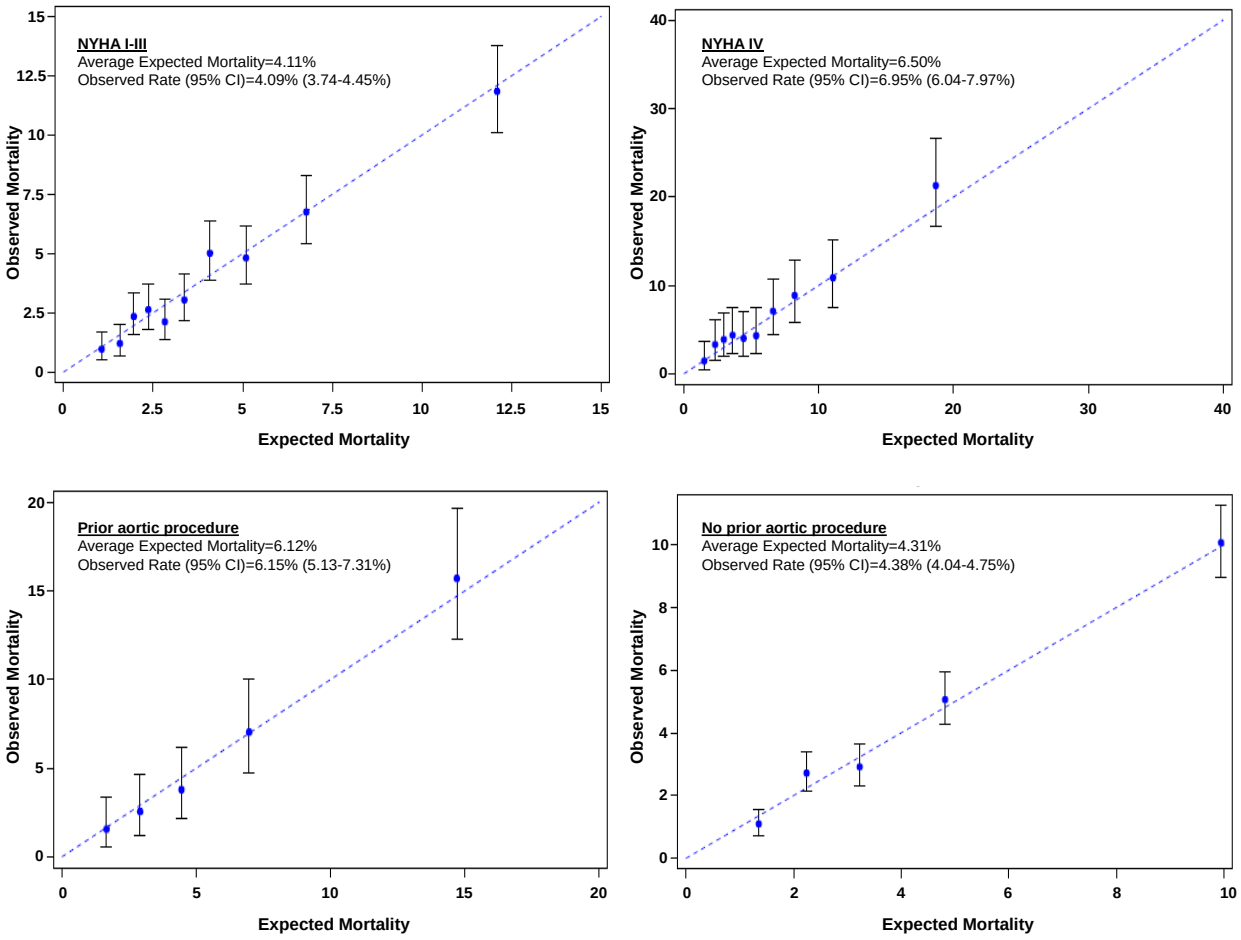
missing data, as they will not benefit from full risk adjustment (thereby encouraging more complete data collection). We elected not to use multiple imputation for this purpose, because of: 1) a lack of standard formulas to calculate hospital-specific risk-adjusted mortality rates and 95% confidence intervals (CIs) when combining multiple imputation with hierarchical modeling empirical Bayes estimators; 2) the computational burden of using multiple imputation in production runs of the TVT Registry feedback report; and 3) the intended slight negative bias to encourage complete data collection, as described earlier.

**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*)

We acknowledge that the interpretation of hospital-specific results are impacted by the exclusion of hospitals with insufficient data completeness. In particular, a hospital's odds ratio must be interpreted as a comparison between the hospital's results and the set of TVT centers that were included in the analysis. Although such a comparison is less desirable than a comparison between a hospital and all US TAVR centers, it is still internally valid as a comparison with the centers included. Importantly, the model will be re-estimated for the TVT registry participant feedback report each quarter. Future analyses for the TVT feedback report will likely include a higher proportion of TAVR centers as the proportion of sites with complete data increases.

**Supplemental Figure 1. Calibration of 30-day mortality risk adjustment model in key subgroups.** Deciles (except for prior aortic procedure, which is presented in quintiles due to small numbers) of predicted odds are plotted against the observed odds of death at 30 days after TAVR, with 95% confidence intervals. The dashed line represents perfect calibration.






## 3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

**3a. Byproduct of Care Processes**

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

**3a.1. Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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 clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

**3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

**3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**

**Attachment:**

**3c. Data Collection Strategy**

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

**3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

**IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.**

Availability:

All hospitals performing TAVR participate in the STS/ACC TVT Registry as a condition of a CMS coverage with evidence decision. Hospitals report patient demographics, medical history and risk factors, frailty and health status, hospital presentation, procedural details, medications, laboratory values, in-hospital complications, and 30-day and 1-year follow-up to the registry. Except for the Kansas City Cardiomyopathy Questionnaire and six minute walk test, all data elements required for this measure are routinely assessed, available in the medical record and acquired during the delivery of standard cardiac care to this patient population. In addition, abstractors capture data in the follow-up assessment period by obtaining medical records from physicians after the patient is discharged. Institutions abstract and submit data manually using a web-based tool via a secure web-portal as part of the site´s enrollment in the Registry.

Sampling:

No sampling is permitted. The facility contract for participation in the STS/ACC TVT Registry requires that all patients treated with TAVR at all hospitals must be included. There will be no discrimination or bias with respect to inclusion on the basis of sex, race, or religion.

Patient confidentiality:

Patient confidentiality is preserved as the data are in aggregate form. The TVT Registry dataset, comprised of approximately 300 data elements, was created by a panel of experts using available ACC-AHA and STS guidelines, existing registries, clinical trials, and other evidentiary sources. Private health information (PHI), such as social security number, is collected. The intent for collection of PHI is to allow for registry

interoperability, for the generation of patient-level drill downs in the Outcomes Reports, as well as providing data to CMS as part of the TAVR "coverage with evidence determination". When using the NCDR web-based data collection tool, direct identifiers are entered but a partition between the data collection process and the data warehouse maintains the direct identifiers separate from the analysis datasets. The minimum level of PHI is transmitted to other stakeholders, meeting the definition of a Deidentified, or Limited Dataset as such term is defined by the Health Insurance Portability and Accountability Act of 1996.

Data collection within the STS/ACC TVT Registry conforms to laws regarding protected health information. Physician and/or institutional confidentiality are maintained by de-identified dashboard reports. There is no added procedural risk to patients through involvement in the TVT Registry. No testing, time, risk, or procedures beyond those required for routine care will be imposed. The primary risk associated with this measure is the potential for a breach of patient confidentiality. The STS and ACCF have established a robust plan for ensuring appropriate and commercially reasonable physical, technical, and administrative safeguards are in place to mitigate such risks.

Data are maintained on secure servers with appropriate safeguards in place at the ACCF. The project team periodically reviews all activities involving protected health information to ensure that such safeguards including standard operating procedures are being followed. The procedure for notifying the STS and ACCF of any breach of confidentiality and immediate mitigation standards is communicated to participants. STS and ACCF limits access to Protected Health Information (PHI), and to equipment, systems, and networks that contain, transmit, process or store PHI, to employees who need to access the PHI for purposes of performing STS and ACCF´s obligations to participants who are in a contractual relationship with the ACCF. All PHI are stored in a secure facility or secure area within ACCF´s facilities, which has separate physical controls to limit access, such as locks or physical tokens.

The secured areas are monitored 24 hours per day, 7 days per week, either by employees or agents of ACCF by video surveillance, or by intrusion detection systems.

Each participant who has access to the STS/ACC TVT Registry website must have a unique identifier. The password protected webpages have implemented inactivity time-outs. Encryption of wireless network data transmission and authentication of wireless devices containing each participant´s information ACCF´s network is required. PHI may only be transmitted off of ACCF´s premises to approved parties, which shall mean: A subcontractor who has agreed to be bound by the terms of the Business Associate Agreement between the STS, ACCF and the TVT Registry Participant.

Time of Data collection: 1 full time employee can enter on average roughly 1,200 patient records per year (citation: ACC Marketing Intelligence Team)

**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)*.

The STS/ACCF TVT Registry provides evidence-based solutions for cardiac surgeons, cardiologists and other medical professionals committed to excellence in cardiovascular care. TVT Registry hospital participants receive confidential benchmark reports that include access to measure specifications, the eligible patient population, exclusions, and model variables (when applicable). In addition to hospitals, NCDR Analytic and Reporting Services provides consenting hospitals' aggregated data reports to interested federal and state regulatory agencies (e.g. CMS), multi-system provider groups, third-party payers, industry partners and other organizations that have an identified quality improvement initiative, or post approval study that supports TVT Registry participating facilities. Lastly, the Registry also allows for licensing of the measure specifications outside of the Registry.

Measures that are aggregated and submitted to NQF are intended for public reporting and therefore there is no charge for a standard export package. However, on a case by case basis, requests for modifications to the standard export package will be available for a separate charge.

## 4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

**4a. Accountability and Transparency**

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

**4.1. Current and Planned Use**

*NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.*

| Specific Plan for Use | Current Use (for current use provide URL) |
|---|---|
| Public Reporting | Regulatory and Accreditation Programs<br>CMS CED<br>https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=293&bc=ACAAAAAAQAAA&<br>Transcatheter Valve Certification<br>https://cvquality.acc.org/accreditation/services/TCV<br>Quality Improvement (external benchmarking to organizations)<br>STS/ACC TVT Registry™<br>https://www.ncdr.com/WebNCDR/tvt/publicpage/home |

**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

The TVT Registry is sponsored by the STS and ACC. It is a national quality improvement registry intended to improve the quality of care provided to patients receiving transcatheter valve replacement and repair procedures since its inception in 2011. It provides a streamlined, consolidated method of collecting, monitoring and reporting clinically relevant data within a framework that ensures both hospital and patient confidentiality. This enables participants to better focus on ACC/AHA guideline-recommended care and to develop new ways for the registry to advance improvements in care and examine newer clinical questions. There are 656 participating sites across the United States with 249,022 cumulative records as of October 16, 2019.

**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?*)

- The measure is currently published in STS/ACC TVT Registry Outcomes Reports for quality improvement and benchmarking.
- It is included in the Transcatheter Valve Certification. Facilities have to have a process to review the metrics, including 30-day risk-adjusted mortality, at least quarterly with the defined multidisciplinary team. Any trend in suboptimal outcomes require action plans, case reviews or root cause analysis that have to be reviewed by the accreditation team. It must identify corrective actions and initiatives to improve the composite measures.
- STS and ACC is currently developing a plan to publicly report site outcomes in 2021.

**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.*)

STS and ACC is currently developing a plan to publicly report site outcomes in 2021.

**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.**

Performance results are distributed to all TVT registry participants as part of quarterly benchmark reports, which provide a detailed analysis of an institution ´s individual performance in comparison to the entire registry population from participating hospitals across the nation. Reports include an executive summary dashboard, at-a-glance assessments, and patient level drill-downs. Registry participants also have access to companion guide for the outcomes report as well as for all risk models. These provide common definitions and detailed specifications to assist with interpretation of the model and of performance rates.

**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.**

Results are provided as part of quarterly performance report, which includes a rolling 3 years of data.

Participating hospitals in the TVT registry report on the following: patient demographics; provider and facility characteristics; adverse event rates; TVT performance measures and select quality measures and outcomes and compliance with STS and ACC/AHA clinical guideline recommendations.

The majority of the required data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions manually capture data using a free web-based tool. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

There are a number of methods used to educate and provide general support to registry participants. This includes the following:

- Registry Site Manager Calls are available for all TVT Registry participants. RSM calls are provided as a source of communication between the TVT Registry and participants to provide education, Registry updates and a live chat Q and A session on a monthly basis.
- A Registry Site Manager Call was devoted to this measure when it was first released in outcome reports. Measure developers provided a presentation.
- New User Calls are available for TVT Registry participants and are intended for assisting new users with their questions. ?
- NCDR and STS Adult Cardiac Surgery Database Annual Conferences.

These annual conferences are well-attended and energetic program at which participants from across the country come together to hear about new STS or NCDR initiatives as well as registry-specific updates. During informative general sessions, attendees can learn about topics such as transcatheter therapies, report dashboards, risk models, data quality and validation, and value-based purchasing.

- Release notes (for outcomes reports)
- Risk model and outcome report companion guides.

- Clinical Support -The Registry Support and Clinical Quality Consultant Teams are available to assist participating sites with questions Monday through Friday, 9:00 a.m. - 5:00 p.m. ET.

**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.**

Initially, feedback on measures are obtained during an "open comment period" during the development of the measure in 2017. That feedback is reviewed by the measure developers to consider modifications prior to finalizing the model.

Measure developers provided a webinar to Registry Participants and other stakeholders when the measure was first released in the outcome reports in 2018. Feedback was obtained during a question and answer session after the presentation.

Feedback is also obtained through monthly registry site manager monthly calls, ad hoc phone calls tracked with Salesforce software, and during registry –specific break-out sessions at the NCDR's annual meeting. Registry Steering Committee members may also provide feedback during regularly scheduled calls.

**4a2.2.2. Summarize the feedback obtained from those being measured.**

In both the open comment period and webinar for this measure, the feedback was generally positive. There were no critical comments that led to a change in the measure.

**4a2.2.3. Summarize the feedback obtained from other users**

No other feedback was received.

**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.**

No changes have been made to the measure.

**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.**

Between 2014 and 2017, the aggregate 30-day TAVR mortality rate in the analysis population decreased from 5.9% to 2.7%, representing a relative decrease of 54%.

Overall 30-day Mortality:

2014: 5.9%

2015: 4.2%

2016: 3.1%

2017:2.7%

We estimate that the improvement in mortality over time is partially explained by a shift in case mix toward lower risk patients. It may also demonstrate improvements in at the facility level. In the hierarchical logistic regression model for the time period June 2013 to May 2016 accounting for differences in case mix, the

estimated odds of mortality decreased 15% per year (odds ratio per year 0.85, 95% CI 0.78 to 0.93, p<0.001), which is a more appropriate estimate of improvements in care at a facility level. We will continue to monitor changes in performance scores and the underlying variables which may impact these changes 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.**

There were no unintended consequences to individuals or populations identified during testing or implementation.

**4b2.2. Please explain any unexpected benefits from implementation of this measure.**

Sites have reported being able to develop process improvement mechanisms and improve their documentation practices as a result of the TVT Registry implementing this measure at the dashboard level which provides a patient drill down feature that helps analyze the sites performance at a granular level.

## 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.**

#2561: STS Aortic Valve Replacement (AVR) Composite Score

**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.**

While this measure focuses on a different population (ie those undergoing surgical AVR) and different outcomes, the current measure has been harmonized to the extent possible. Residual differences in the two models include the following: 1. Some variables are unique to each population/procedure/measure (e.g. TAVR 30-day RAM includes variables unique to the procedure such as gait speed, KCCQ, access site, porcelain aorta and aortic valve morphology). 2. The outcome of each measure is different. TAVR 30-day RAM is subset of the STS AVR Composite Score (which includes 30-day mortality as well as 5 morbidities). 3. The patient

population of each measure is different. TAVR 30 day RAM is only patients who had a transcatheter aortic valve replacement procedures. STS AVR Composite is for all patients having an aortic valve replacement (which MAY include a TAVR).

**5b. Competing Measures**

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
**OR**
Multiple measures are justified.

**5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

**Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)**

N/A

## Appendix

**A.1 Supplemental materials may be provided in an appendix.** All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment **Attachment:** Arnold_30_Day_Mortality_after_TAVR.pdf

## Contact Information

**Co.1 Measure Steward (Intellectual Property Owner):** American College of Cardiology

**Co.2 Point of Contact:** Susan, Fitzgerald, sfitzger@acc.org, 240-620-5444-

**Co.3 Measure Developer if different from Measure Steward:** American College of Cardiology

**Co.4 Point of Contact:** Susan, Fitzgerald, sfitzger@acc.org, 240-620-5444-

## 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.**

STS/ACC TVT Registry Risk Model Workgroup:

Fred H. Edwards, MD7 – workgroup chair

STS Workgroup Members: David M. Shahian MD3; Vinod H. Thourani MD6

ACC Workgroup Members: Suzanne V. Arnold MD MHA1; David J. Cohen MD MSc1; Eric D. Peterson MD MPH2

CMS Representative: Rosemarie Hakim PhD4

Data analytic center staff: Sreekanth Vemulapalli MD2; Amanda Stebbins MS2; J. Matthew Brennan MD MPH2, Sean M. O'Brien, PhD2

1Saint Luke's Mid America Heart Institute and University of Missouri-Kansas City, Kansas City, MO; 2Duke University, Durham, NC;

3Lahey Hospital and Medical Center and Harvard Clinical Research Institute, Boston, MA;

4Centers for Medicare & Medicaid Services, Baltimore, MD;

5University of Colorado School of Medicine, Aurora, CO;

6Medstar Washington Hospital Center/Georgetown University, Washington, DC;

7University of Florida College of Medicine-Jacksonville, Jacksonville, FL

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 2018

**Ad.3 Month and Year of most recent revision:**

**Ad.4 What is your frequency for review/update of this measure?** No formal review scheduled at this time.

**Ad.5 When is the next scheduled review/update for this measure?**

**Ad.6 Copyright statement:** American College of Cardiology Foundation All Rights Reserved

**Ad.7 Disclaimers:** STS and ACC do not have a web page dedicated to the TVT Registry measure specification. Participants can access a risk model companion guide to help them understand the model. The manuscript is also a publicly available resource.

**Ad.8 Additional Information/Comments:** STS and ACC appreciate the opportunity to submit measures for this NQF endorsement maintenance project.

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 #:** 0965

**Corresponding Measures:**

**De.2. Measure Title:** Discharge Medications (ACE/ARB and beta blockers) in Eligible ICD/CRT-D Implant Patients

**Co.1.1. Measure Steward:** American College of Cardiology

**De.3. Brief Description of Measure:** Proportion of patients undergoing ICD/CRT-D implant who received prescriptions for all medications (ACE/ARB and beta blockers) for which they are eligible at discharge.

**1b.1. Developer Rationale:** This measure is intended to assess the extent to which eligible patients receive evidence-based medications that are indicated at hospital discharge following ICD placement. This measure focuses on processes of care that are supported by guidelines for optimal care for patients undergoing ICD placement.

**S.4. Numerator Statement:** Generator patients who receive all medications for which they are eligible:

1. ACE/ARB prescribed at discharge (if eligible for ACE/ARB as described in denominator) AND
2. Beta blockers prescribed at discharge (if eligible for beta blockers as described in denominator)

**S.6. Denominator Statement:** All generator patients surviving hospitalization who are eligible to receive either an ACE/ARB or beta blocker at discharge.

**S.8. Denominator Exclusions:** None

**De.1. Measure Type:** Composite

**S.17. Data Source:** Registry Data

**S.20. Level of Analysis:** Facility

**IF Endorsement Maintenance – Original Endorsement Date:** Feb 05, 2013 **Most Recent Endorsement Date:** Feb 19, 2016

## Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

## Criteria 1: Importance to Measure and Report

### 1a. Evidence

**Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.**

**1a. Evidence.** The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☐ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☐ **No**
- **Evidence graded?** ☐ **Yes** ☐ **No**

**Summary of prior review in 2016**

- This composite measure has two process measure components. Each component measure should must meet the evidence criterion.
- This composite measure has two component measures that assess if all patients with an ICD implant surviving hospitalization receive all medications (ACE/ARB and beta blockers) for which they are eligible at discharge. Because the beta-blocker component may be applied to two separate patient populations (patients with previous MI and patients with LVSD), the developer has provided evidence supporting the use of beta blockers in each of these populations separately. The developer provides diagrams demonstrating how receiving beta-blockers for a previous MI, LVSD and ACEI/ARBs for LVSD are linked to patient outcomes.

Beta-blocker for previous MI

- The developer provided four guidelines with six guideline statements that recommend beta-blocker therapy for patients with HF or prior MI. Of the six guideline statements, all are Class I recommendations with two A, three B, and one C level of evidence.
- One prospective cohort study and one meta-analysis were published after the publication of the 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline. The analysis concluded that the use of beta-blockers in patients with stable CAD was associated with a lower risk of cardiovascular mortality.
- Process measure (Box 3) → Based on systematic review (Box 4) → No QQC (moderate is highest rating) (Box 6) → Class I, all but one are Grade A or B → Moderate rating

Beta-blocker for LVSD

- The developer provided two guidelines with four guideline statements that recommend beta-blocker therapy for patients with LVSD, with or without prior MI. Of the four guideline statements, all are Class I recommendations with two A, one B, and one C level of evidence.
- New for this review cycle: the developer provides one additional guideline with one guideline statement focused on management of patients with ventricular rhythm arrhythmias and the prevention of sudden cardiac death. The guideline statement is a Class I recommendation with an A level of evidence.
- One RCT, one prospective cohort study, and two meta-analyses were published after the publication of the 2013 ACCF/AHA Guideline for the Management of Heart Failure. The analysis concluded that the use of beta-blockers in patients with stable CAD was associated with a lower risk of cardiovascular mortality.
- Process measure (Box 3) → Based on systematic review (Box 4) → No QQC (moderate is highest rating) (Box 6) → Class I, all but one are Grade A or B → Moderate rating

ACE/ARBs for LVSD

- The developer provided two guidelines with four guideline statements that recommend ACE/ARBs for patients with LVSD, with or without prior MI. Of the four guideline statements, all are Class I recommendations, and all are an A level of evidence.
- New for this review cycle:
    - The 2013 ACCF/AHA Guideline for the Management of Heart Failure had a focused update in 2017. The updated guideline has four guideline statements recommending ACE/ARBs either alone or with beta blockers for patients with chronic HF (Class I, Level of Evidence A). For some patients, replacing the ACE/ARB with an ARNI is recommended (Class I, Level of Evidence B-R).
    - The developer provides one additional guideline with one guideline statement focused on management of patients with ventricular rhythm arrhythmias and the prevention of sudden cardiac death. The guideline statement is a Class I recommendation with an A level of evidence.
- One meta-analysis was published after the publication of the 2013 ACCF/AHA Guideline for the Management of Heart Failure.
- Process measure (Box 3) → Based on systematic review (Box 4) → No QQC (moderate is highest rating) (Box 6) → Class I, all are Grade A → Moderate rating

**Changes to evidence from last review**

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure as indicated in the summary above.**

***Questions for the Committee:***

- The evidence provided by the developer is updated, and directionally the same, compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?

**Guidance from the Evidence Algorithm**

- For each component measure: Process measure (Box 3) → Based on systematic review (Box 4) → No QQC (moderate is highest rating) (Box 6) → Class I, all are Grade A or B → Moderate rating

**Preliminary rating for evidence:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

## 1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

**Maintenance measures – increased emphasis on gap and variation**

**1b. Performance Gap.** The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- Performance data at the composite level show a moderate opportunity for improvement and demonstrate improvement since the last review cycle.
- Performance data are not provided for the individual measure components.

Summary of performance data from 2015 review cycle:

In 2011-2012 a total of 243,186 patients at 1552 hospitals were analyzed and 195,563 patients at 1606 hospitals in 2013-14. Data from 2011-12 indicated a mean of 74% and 50th percentile results at 76%. Data from 2013-14 indicated a mean of 78% and 50th percentile results at 79%.

Updated performance data:

Performance scores from the National Cardiovascular Data Registry's ICD Registry, a national quality improvement registry are provided below. Scores are from the U.S. hospitals participating in the registry.

| Year | N | Mean | Std Dev | 0% (Min) | 5% | 10% | 25% | 50% (Med) | 75% | 90% | 95% | 100% (Max) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2017 | 1674 | 83% | 18% | 0% | 48% | 60% | 74% | 88% | 97% | 100% | 100% | 100% |
| 2018 | 1574 | 83% | 20% | 0% | 47% | 60% | 75% | 88% | 97% | 100% | 100% | 100% |

N = number of facilities, results rounded to whole numbers for ease of display in table

**Disparities**

- Mean scores appear similar across groups and when compared to overall mean. Some variation at the median, with Hispanic, Black, and Other groups showing higher results than White and Non-White groups. Dual Eligible group scores are very similar to overall scores.

Performance scores from the National Cardiovascular Data Registry's ICD Registry. Scores stratified by gender, age, race/ethnicity, and dual eligibility provided from U.S. hospitals participating in the registry.

Measurement Year 2017

| | N | Mean | Std Dev | 0% (Min) | 5% | 10% | 25% | 50% (Med) | 75% | 90% | 95% | 100% (Max) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Male | 1674 | 83% | 18% | 0% | 50% | 60% | 75% | 88% | 97% | 100% | 100% | 100% |
| Female | 1674 | 83% | 21% | 0% | 40% | 50% | 75% | 90% | 100% | 100% | 100% | 100% |
| Age <65 | 1674 | 84% | 20% | 0% | 50% | 58% | 77% | 91% | 100% | 100% | 100% | 100% |
| Age =>65 | 1674 | 82% | 19% | 0% | 44% | 59% | 73% | 87% | 97% | 100% | 100% | 100% |
| Hispanic | 1674 | 84% | 27% | 0% | 0% | 50% | 76% | 100% | 100% | 100% | 100% | 100% |
| White non-Hispanic | 1674 | 83% | 18% | 0% | 50% | 60% | 74% | 88% | 97% | 100% | 100% | 100% |
| Black non-Hispanic | 1674 | 84% | 24% | 0% | 33% | 50% | 76% | 97% | 100% | 100% | 100% | 100% |
| Other | 1674 | 86% | 27% | 0% | 0% | 50% | 86% | 100% | 100% | 100% | 100% | 100% |
| Non-White | 1674 | 82% | 19% | 0% | 47% | 57% | 74% | 87% | 97% | 100% | 100% | 100% |
| Dual Eligibility | 1674 | 83% | 16% | 0% | 50% | 62% | 74% | 87% | 96% | 100% | 100% | 100% |

N = number of facilities, results rounded to whole numbers for ease of display in table

Measurement Year 2018

| | N | Mean | Std Dev | 0% (Min) | 5% | 10% | 25% | 50% (Med) | 75% | 90% | 95% | 100% (Max) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Male | 1574 | 83% | 20% | 0% | 48% | 58% | 75% | 89% | 98% | 100% | 100% | 100% |
| Female | 1574 | 82% | 23% | 0% | 33% | 53% | 75% | 91% | 100% | 100% | 100% | 100% |
| Age <65 | 1574 | 85% | 21% | 0% | 50% | 60% | 78% | 93% | 100% | 100% | 100% | 100% |
| Age =>65 | 1574 | 82% | 21% | 0% | 43% | 55% | 72% | 88% | 98% | 100% | 100% | 100% |
| Hispanic | 1574 | 85% | 26% | 0% | 0% | 50% | 77% | 100% | 100% | 100% | 100% | 100% |
| White non-Hispanic | 1574 | 83% | 20% | 0% | 44% | 58% | 74% | 89% | 99% | 100% | 100% | 100% |
| Black non-Hispanic | 1574 | 85% | 25% | 0% | 33% | 50% | 78% | 100% | 100% | 100% | 100% | 100% |
| Other | 1574 | 86% | 28% | 0% | 0% | 50% | 86% | 100% | 100% | 100% | 100% | 100% |
| Non-White | 1574 | 82% | 20% | 0% | 42% | 56% | 75% | 88% | 97% | 100% | 100% | 100% |
| Dual Eligibility | 1574 | 83% | 16% | 0% | 53% | 62% | 75% | 86% | 96% | 100% | 100% | 100% |

N = number of facilities, results rounded to whole numbers for ease of display in table

***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**

**1c. Composite –** Quality Construct and Rationale

**Maintenance measures – same emphasis on quality construct and rationale as for new measures.**

**1c. Composite Quality Construct and Rationale**. The quality construct and rationale should be explicitly articulated and logical; a description of how the aggregation and weighting of the components is consistent with the quality construct and rationale also should be explicitly articulated and logical.

- This composite is an all-or-none composite. Patients must receive all medications for which they're eligible in order to be included in the numerator.
- The developer states the all-or-none composite reflects the strong recommendations for each process of care included in the composite. They state that combining the measures into one composite provides a perspective of the overall quality of medical therapy while reducing information burden.
- The developers state they conducted empirical analyses exploring the relationship between performance on the composite and clinical outcomes and discovered:
    - Patients discharged on appropriate medical therapy were less likely to experience adverse outcomes
    - Fewer patients treated at hospitals performing well on the composite experienced adverse outcomes compared with those treated at lower performing hospitals
    - Outcomes studied included readmission and mortality at 6 months.

***Questions for the Committee:***

- Are the quality construct and a rationale for the composite explicitly stated and logical?
- Is the method for aggregation and weighting of the components explicitly stated and logical?

**Preliminary rating for composite quality construct and rationale:**
**☐ High ☒ Moderate ☐ Low ☐ Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)**

1a.

- N/A
- Data fits well when discussing the medications for LVSD or MI; however, I am having trouble connecting the ICD/CRT implant into the discussion/rationale/data. More information here may be helpful.
- There is strong evidence that ACE/ARB and beta-blockers have net clinical benefit in patients after MI and with HFrEF, as documented. However, the presented evidence does not speak to patients with an ICD specifically, so the justification for the measure focus is not clear. Also, the newr class of vasodilators (ANRI) should be added. Lastly, the evidence is for *treatment* with these drugs, not for a one-time prescription as operationalized in the measure.
- Evidence is solid

- Closely relates to outcome being measured. Not aware of any new studies/information that changes the evidence for this measure.

1b.

- Overall less than optimal with variation. Not much disparity by race, gender, age etc
- A gap was show and still warrants more progress.
- Sufficient evidence to show a performance gap and variability as well as disparities.
- Yes a gap still exists
- Consider discussion; uncertain about the comment "fewer patients experienced adverse events than those with lower performing hospitals." Uncertain how, if it does, correlate to minority population and those treated at "lower performing hospitals"

1c.

- Yes
- Addressed within the document. No major concerns.
- composite formation is reasonable
- yes
- Prior comment relates here; consider discussion of provider-specific performance for consistency in performance

## Criteria 2: Scientific Acceptability of Measure Properties

**2a. Reliability: Specifications and Testing**

**2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data**

**2c. For composite measures: empirical analysis support composite approach**

### Reliability

**2a1. Specifications** requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

**2a2. Reliability testing** demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

### Validity

**2b2. Validity testing** should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

**2b2-2b6. Potential threats to validity** should be assessed/addressed.

**Composite measures only:**

**2d. Empirical analysis to support composite construction**. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

**Complex measure evaluated by Scientific Methods Panel**? ☐ **Yes** ☒ **No**

**Evaluators:** NQF Staff

Scientific Acceptability: Preliminary Analysis Form

Reliability

- Data element: There were no updates from the previous submission. A sample of 627 patients from 25 hospitals were selected for inter-rater reliability (IRR) of the extracted data elements. This was performed by an independent contractor.
    - 2015 submission included IRR for six data elements. Kappa values ranged from 0.33 (LVEF assessed) to 0.96 (Procedure type) with most values > 0.60. A kappa > 0.70 is considered acceptable inter-rater reliability. This IRR was performed on data from 2010.
    - 2015 submission indicates the IRR of data elements is conducted on a 3-year rotating cycle and that the elements for this measure will be reviewed during the upcoming audit process.
- Score-level: Developers used a split-sample methodology. The cohort was split into two random samples and scores calculated using the same timeframe. For the performance rates and social risk data, unadjusted rates were calculated, and a Pearson correlation coefficient and ICC were computed.
    - For 2017, Pearson correlation coefficient: 0.59, ICC: 0.82, indicating a moderate to strong reliability
    - For 2018, Pearson correlation coefficient: 0.52, ICC: 0.79, indicating a moderate to strong reliability

Validity

No changes from 2015 submission. From the 2015 preliminary analysis:

- Empiric testing was conducted at the level of the data element and measure score using 93,971 Medicare FFS patients who were at least 65 years of age and underwent ICD implantation in 2010 or 2011.
- The analyses included the association of patient and hospital performance on the composite measure with adverse outcomes, specifically mortality and readmission at 6 months following hospital discharge and the association between hospital-level performance on the measure and the combination of mortality or readmission at 6 months. The developer provides patient-level and hospital level results:
    - A significantly smaller proportion of patients discharged on the appropriate medical therapy died or were readmitted within 6 months of hospital discharge (without meds = 28.37% vs. with meds = 36.28%).

- Patients treated at hospitals that performed better on the measure had better unadjusted outcomes that those treated at hospitals that performed worse on the measure (correlation coefficient (-0.0998), $p<0.001$).

Composite

The developer conducted empirical validity analysis of the relationship between the individual component measures and the overall composite measure. The individual components were strongly correlated (0.70 or higher for all analyses) with the overall composite. A logistic regression analysis provided by the developer demonstrates that the ACE/ARB and Beta Blocker measures explained 89.0% and 68.0% of the overall variance, respectively.

***Questions for the Committee regarding reliability:***

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The staff is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss or vote on reliability?

***Questions for the Committee regarding validity:***

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss or vote on validity?

***Questions for the Committee regarding composite construction:***

- Do you have any concerns regarding the composite construction approach (e.g., do the component measures fit the quality construct and add value to the overall composite? Are the aggregation and weighting rules consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible?)?
- The staff is satisfied with the composite construction. Does the Committee think there is a need to discuss or vote on the composite construction approach?

**Preliminary rating for reliability:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

**Preliminary rating for validity:** ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

**Preliminary rating for composite construction:** ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)**

2a1.

- There is some difficulty with element reliability but score-level is acceptable
- No concerns with reliability.

- Reliability is adequately assessed and the data are derived from a registry with regular audits of data quality.
- satisfactory, would update to include ARNI in ARB group
- Concur with staff recommendations; no new data or information recommended.

2a2.

- No
- This isn't my area of expertise, but seems good.
- No
- Low IRR regarding EF is concerning, given the importance of EF on this outcome and procedure.
- Concur with staff recommendation as stated.

2b1.

- no concerns
- This isn't my area of expertise, but seems good.
- Convincing evidence for validity with the association of prescribing rates and outcomes
- Better validity demonstration than most measures
- Concur with Staff recommendation as stated.

2b4-7.

- I don't see any significant threats to validity
- No concerns.
- no concerns
- Would like to see data that there is no trend to higher missing data for shorter length of stay patients, this could introduce systemic bias with under-representation of lower risk patients
- No concerns about threats to validity.

2b2-3.

- Not risk adjusted
- The measure specifications, which include exclusion of medications due to contraindication address the exclusions. These are considered in the numerator, which seems fine.
- no concerns
- not evaluated
- Insufficient knowledge

2c.

- performs well
- No concerns
- It is a fairly simple composite that just says the patient must receive all indicated meds to pass. One could as easily call this a regular process measure with several components. no concerns.
- Yes
- Yes, analysis demonstrate component measures

## Criterion 3. Feasibility

**Maintenance measures – no change in emphasis – implementation issues may be more prominent**

**3. Feasibility** is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer states that all data elements associated with this measure are routinely generated and acquired during the delivery of standard cardiac care to this patient population. They state most data elements exist in a structured format within an EHR and that data can extracted electronically.
- The developer states a full-time employee can enter roughly 1,200 patient records per year on average.
- The developer notes that participation in the registry is a requirement for Medicare reimbursement purposes and that almost all hospitals that implant ICDs already participate for this reason.

***Questions for the Committee:***

- Do you agree that the required data elements are routinely generated and used during care delivery?
- Is the data collection strategy feasible?

**Preliminary rating for feasibility:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 3: Feasibility**

- The measure is feasible for participating hospitals
- No concerns
- Those are routinely collected data fields
- mandated registry based
- Data collection is feasible and participation and data collection required for Medicare reimbursement; high motivation for compliance and clearly stated how it will be performed. One possible concern is the incompatibility between EHR systems, manual record keeping in patients who have long-standing CVD, have seen multiple providers, sites of care, etc.

## Criterion 4: Usability and Use

**Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences**

## 4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

**4a. Use** evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

**4a.1. Accountability and Transparency.** Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

**Current uses of the measure**

**Publicly reported?** ☒ **Yes** ☐ **No**

**Current use in an accountability program?** ☐ **Yes** ☒ **No** ☐ **UNCLEAR**

**OR**

**Planned use in an accountability program?** ☐ **Yes** ☐ **No**

**Accountability program details**

Public Reporting

NCDR Public Reporting

**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

- The developer states that each participant receives quarterly feedback reports providing a detailed analysis of the participant's performance including benchmarking. Participants also have access to a guide to help interpret performance results.

**Feedback on the measure by those being measured or others**

- The developer reports that feedback is typically obtained through monthly registry site manager calls, ad hoc calls, and break-out sessions at the registry's annual meeting. They report feedback has generally been supportive and positive regarding the measure and registry. They report clarifying the language and adding CRT-D implant patients in response to feedback.

**Additional Feedback:**

***Questions for the Committee**:*

- How have the performance results been used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

**Preliminary rating for Use:** ☒ **Pass** ☐ **No Pass**

## 4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

**4b. Usability** evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

**4b.1 Improvement.** Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

**Improvement results**

The mean rate of performance has improved over time from 74% when the measure was first released (2011-12) to 78% in 2013-14 and 83% in the most recent data year (2018).

**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 reports no unexpected findings or unintended consequences.

**Potential harms**

None noted.

**Additional Feedback:**

***Questions for the Committee**:*

- Are you aware of any unintended consequences related to this measure?
- Do the benefits of the measure outweigh any potential unintended consequences?

**Preliminary rating for Usability and use:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

**Committee Pre-evaluation Comments:**
**Criteria 4: Usability and Use**

4a1.

- It is not publicly reported but is part of the NCDR ICD registry
- Results are publicly reported.
- Developer states that the measure is publicly reported but not used in an accountability program.
- useful and higher rates demonstrated over time
- Data are primarily reported back to participants in the registry through reports and assistance with analyzing reports, annual conferences, etc. The plan seems reasonable and is underway. More creative dissemination could be explored for educating consumers, patient organizations, etc. consider the frequent use of ICD implants and the growing knowledge patients have of treatment choices. Yes, feedback is incorporated into the measure routinely.

4b1.

- No significant harms identified
- Positive trend over time (74-83% performance).
- Hospitals could use the information to educate clinicians on prescribing. No concerns about unintended consequences.
- no obvious harms
- Demonstrated improvement in mean performance of measure demonstrating its familiarity with intended users. Benefits of the measure outweigh unintended consequences (none, however, are noted).

## Criterion 5: Related and Competing Measures

**Related or competing measures**

0066: Coronary Artery Disease (CAD): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy - Diabetes or Left Ventricular Systolic Dysfunction (LVEF < 40%)
0070: Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)
0070e: Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)
0071: Persistence of Beta-Blocker Treatment After a Heart Attack
0081: Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)
0081e: Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) or Angiotensin Receptor-Neprilysin Inhibitor (ARNI) Therapy for Left Ventricular Systolic Dysfunction (LVSD)
0083: Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)
0117: Beta Blockade at Discharge
0236: Coronary Artery Bypass Graft (CABG): Preoperative Beta-Blocker in Patients with Isolated CABG Surgery
0594: Post MI: ACE inhibitor or ARB therapy
0696: STS CABG composite score (includes 0236)

**Harmonization**

The developer reports measures are harmonized to the extent possible, except for measure 0081/0081e, which includes ARNIs as an alternative to ACE/ARBs.

**Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures**

- There are several related measures but none the directly compete
- There are many related/competing measures. Stewards state harmonization to the greatest extent as possible. Has the steward considered the ARNI in the value set to account for the ARB use?
- As far as i can tell, the measure is harmonized with the exception of the omission of Angiotensin Receptor-Neprilysin Inhibitor (ARNI) Therapy
- all of the low EF medication measures should incorporate ARNI measurement
- Respondent skipped this question

## Public and Member Comments

**Comments and Member Support/Non-Support Submitted as of: Month/Day/Year**

- **Of the XXX NQF members who have submitted a support/non-support choice:**
  - XX support the measure

- YY do not support the measure

Scientific Acceptability: Preliminary Analysis Form

**Measure Number**: 0965

**Measure Title**: Discharge Medications (ACE/ARB and beta blockers) in Eligible ICD/CRT-D Implant Patients

### Type of Measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☒ **Composite**

## Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☒ **Registry Data**
☐ **Enrollment Data** ☐ **Other**

### Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

## Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

### RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   **Submission document:** "MIF_0965" document, items S.1-S.22

   - The measure's data source is the National Cardiovascular Data Registry (NCDR) ICD Registry with the data dictionary provided. The data elements and the calculation algorithm are described. The developer does not provide ICD-10 codes and specific beta-blockers and ACE/ARBs are not identified.
   - No information is provided on what constitutes an acceptable contraindication.

2. **Briefly summarize any concerns about the measure specifications.**

   - During the previous review cycle for this measure, concerns were raised regarding the lack of codes and specific drug definitions. The notes indicate the measure developer stated that the measure aligned with the guidelines which are at a drug-class level. However, several of the guideline recommendations referenced in the evidence section specifically recommend carvedilol, metoprolol succinate, or bisoprolol for patients with heart failure and reduced ejection fraction.

### RELIABILITY: TESTING

**Submission document:** "MIF_0965" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☒ **Data element** ☐ **Neither**

4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

   **Submission document:** Testing attachment, section 2a2.2

   - Data element: A sample of 627 patients from 25 hospitals were selected for inter-rater reliability (IRR) of the extracted data elements. This was performed by an independent contractor.
   - Score-level: Developers used a split-sample methodology. The cohort was split into two random samples and scores calculated using the same timeframe. For the performance rates and social risk data, unadjusted rates were calculated, and a Pearson correlation coefficient and ICC were computed.

7. **Assess the results of reliability testing**

   **Submission document:** Testing attachment, section 2a2.3

   - Data element
     - 2015 submission included IRR for six data elements. Kappa values ranged from 0.33 (LVEF assessed) to 0.96 (Procedure type) with most values > 0.60. A kappa > 0.70 is considered acceptable inter-rater reliability. This IRR was performed on data from 2010.
     - 2015 submission indicates the IRR of data elements is conducted on a 3-year rotating cycle and that the elements for this measure will be reviewed during the upcoming audit process.
   - Score-level
     - For 2017, Pearson correlation coefficient: 0.59, ICC: 0.82, indicating a moderate to strong reliability
     - For 2018, Pearson correlation coefficient: 0.52, ICC: 0.79, indicating a moderate to strong reliability

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   **Submission document:** Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   **Submission document:** Testing attachment, section 2a2.2

   ☐ **Yes**

☒ **No**

☐ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

Score-level results indicate moderate to high reliability. Data element testing could be stronger with more recent and complete information.

## VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    **Submission document:** Testing attachment, section 2b2.

The measure has been updated to handle denominator inclusions differently. This results in no exclusions.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    **Submission document:** Testing attachment, section 2b4.

    No concerns. While top quartile shows high performance, there is variation and room for improvement in lower quartiles.

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.

    Not applicable.

15. **Please describe any concerns you have regarding missing data.**

    **Submission document:** Testing attachment, section 2b6.

    - No concerns.

- Data submissions to the registry undergo integrity checks that identify missing or inconsistent data. If a submission has excessive missing data or inconsistent data, it is rejected. If a submission passes integrity checks, but fails completeness, it is loaded, but not included in any aggregate calculations. Missing data defaults to "performance not met."
- Auditing program includes a check of billing records against records submitted to the registry to assess completeness of record submissions.

16. **Risk Adjustment – Not applicable to this measure**

16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

16e**. Assess the risk-adjustment approach**

**<u>For cost/resource use measures ONLY</u>:**

17. **Are the specifications in alignment with the stated measure intent?**

☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

## VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**

20. **Method of establishing validity of the measure score:**

☐ **Face validity**

☒ **Empirical validity testing of the measure score**

☐ **N/A (score-level testing not conducted)**

21. **Assess the method(s) for establishing validity**

**Submission document: Testing attachment, section 2b2.2**

No changes from 2015 submission. From the 2015 preliminary analysis:

- Empiric testing was conducted at the level of the data element and measure score using 93,971 Medicare FFS patients who were at least 65 years of age and underwent ICD implantation in 2010 or 2011.

22. **Assess the results(s) for establishing validity**

**Submission document: Testing attachment, section 2b2.3**

No changes from 2015 submission. From the 2015 preliminary analysis:

- The analyses included the association of patient and hospital performance on the composite measure with adverse outcomes, specifically mortality and readmission at 6 months following hospital discharge and the association between hospital-level performance on the measure and the combination of mortality or readmission at 6 months. The developer provides patient-level and hospital level results:
    - A significantly smaller proportion of patients discharged on the appropriate medical therapy died or were readmitted within 6 months of hospital discharge (without meds = 28.37% vs. with meds = 36.28%).
    - Patients treated at hospitals that performed better on the measure had better unadjusted outcomes that those treated at hospitals that performed worse on the measure (correlation coefficient (-0.0998), $p<0.001$).

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

**Submission document:** Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

☐ **Not applicable** (score-level testing was not performed)

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

**Submission document***: Testing attachment, section 2b1.*

☐ **Yes**

☐ **No**

☒ **Not applicable** (data element testing was not performed)

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☐ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

Score-level testing performed demonstrates significant correlation with outcomes of interest. No significant threats to validity.

### FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

☒ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

The developer conducted empirical validity analysis of the relationship between the individual component measures and the overall composite measure. The individual components were strongly correlated (0.70 or higher for all analyses) with the overall composite. A logistic regression analysis provided by the developer demonstrates that the ACE/ARB and Beta Blocker measures explained 89.0% and 68.0% of the overall variance, respectively.

### ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

## Developer Submission

# 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**

0965_evidence_7-1_11.21.2019.docx

**1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?**

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

## 1a. Evidence (subcriterion 1a)

**Component # 1 Beta Blocker Therapy**

**Measure Number** (*if previously endorsed*)**:** 0965

**Measure Title**: Patients with a prior MI and an ICD implant who receive beta blocker therapy at discharge

**IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:** Patients with an ICD implant who receive ACE-I/ARB and beta blocker therapy at discharge

**Date of Submission**: Click here to enter a date

**1a.1.This is a measure of**: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome: Click here to name the health outcome

☐ 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*):

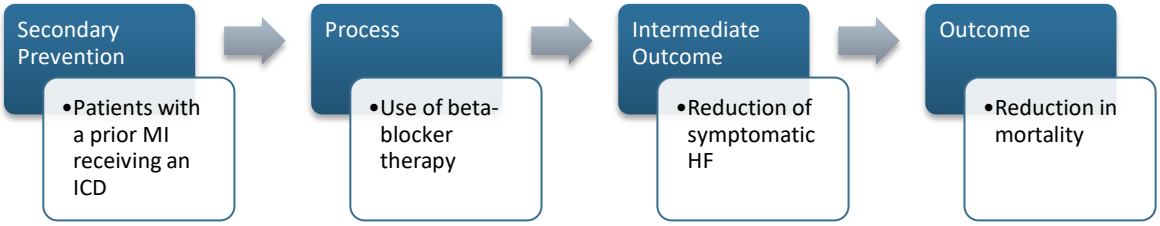
☒ Process: Beta-blocker therapy for patients with a prior MI receiving an ICD

☐ 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.




Beta-blockers reduce morbidity, mortality, and hospitalizations for patients who had a prior myocardial infarction (MI).

**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.**

**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

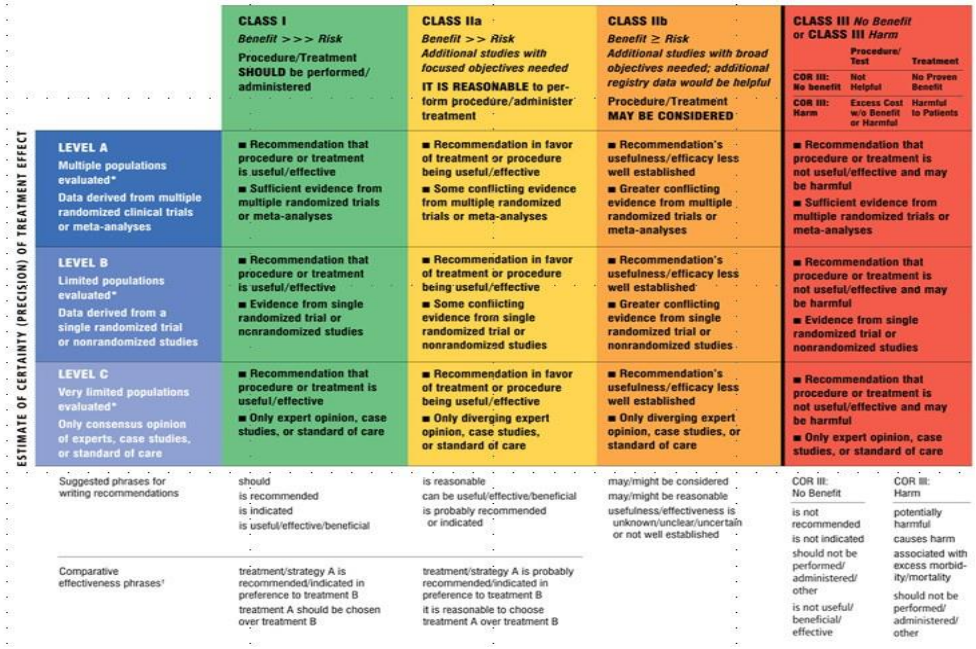
☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

| | |
|---|---|
| **Source of Systematic Review:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | **Amsterdam EA, Wenger NK, Brindis RG, Casey DE Jr, Ganiats TG, Holmes DR Jr, Jaffe AS, Jneid H, Kelly RF, Kontos MC, Levine GN, Liebson PR, Mukherjee D, Peterson ED, Sabatine MS, Smalling RW, Zieman SJ. 2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes: a report of the American College of Cardiology/ American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014;64:e139–228. http://content.onlinejacc.org/article.aspx?articleid=1910086**<br><br>**O'Gara PT, Kushner FG, Ascheim DD, Casey DE Jr, Chung MK, de Lemos JA, Ettinger SM, Fang JC, Fesmire FM, Franklin BA, Granger CB, Krumholz HM, Linderbaum JA, Morrow DA, Newby LK, Ornato JP, Ou N, Radford MJ, Tamis-Holland JE, Tommaso CL, Tracy CM, Woo YJ, Zhao DX. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;61:e78–140, doi:10.1016/j.jacc.2012.11.019. http://content.onlinejacc.org/article.aspx?articleid=1486115**<br><br>**Smith SC Jr., Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, Gibbons RJ, Grundy SM, Hiratzka LF, Jones DW, Lloyd-Jones DM, Minissian M, Mosca L, Peterson ED, Sacco RL, Spertus J, Stein JH, Taubert KA. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation. 2011: published online before print November 3, 2011, 10.1161/CIR.0b013e318235eb4d. http://content.onlinejacc.org/article.aspx?articleid=1147807**<br><br>**Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, Douglas PS, Foody JM, Gerber TC, Hinderliter AL, King SB III, Kligfield PD, Krumholz HM, Kwong RYK, Lim MJ, Linderbaum JA, Mack MJ, Munger MA, Prager RL, Sabik JF, Shaw LJ, Sikkema JD, Smith CR Jr, Smith SC Jr, Spertus JA, Williams SV. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2012;60:e44–164. http://content.onlinejacc.org/article.aspx?articleid=1391404** |

| | |
|---|---|
| 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. | 2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes (p. e 159)<br><br>• In patients with concomitant NSTE-ACS, stabilized HF, and reduced systolic function, it is recommended to continue beta-blocker therapy with 1 of the 3 drugs proven to reduce mortality in patients with HF: sustained-release metoprolol succinate, carvedilol, or bisoprolol. **Class I: Level of Evidence: C**<br><br>2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction (p. e104)<br><br>• Beta blockers should be continued during and after hospitalization for all patients with STEMI and with no contraindications to their use. **Class I: Level of Evidence: B**<br><br>AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update (p. e2435)<br><br>• Beta-blocker therapy should be used in all patients with left ventricular systolic dysfunction (ejection fraction <40%) with heart failure or prior myocardial infarction, unless contraindicated. (Use should be limited to carvedilol, metoprolol succinate, or bisoprolol, which have been shown to reduce mortality.) **Class I: Level of Evidence: A**<br>• Beta-blocker therapy should be started and continued for 3 years in all patients with normal left ventricular function who have had myocardial infarction or ACS. **Class I: Level of Evidence: B**<br><br>2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease (p. e96)<br><br>• Beta-blocker therapy should be started and continued for 3 years in all patients with normal LV function after MI or ACS. **Class I: Level of Evidence: B**<br><br>• Beta-blocker therapy should be used in all patients with LV systolic dysfunction (EF <40%) with heart failure or prior MI, unless contra-indicated. (Use should be limited to carvedilol, metoprolol succinate, or bisoprolol, which have been shown to reduce risk of death.). **Class I: Level of Evidence: A** |
| Grade assigned to the **evidence** | The weight of the evidence in support of most of the recommendations included are rated as Level A, Level B and Level C as noted following each statement. Level |

| | |
|---|---|
| associated with the recommendation with the definition of the grade | A evidence refers to "Data derived from multiple randomized clinical trials or meta-analyses." The weight of the evidence in support of additional recommendations is rated as Level B and C. Level B evidence refers to "Data derived from a single randomized trial, or nonrandomized studies" while Level C evidence refers to "Only consensus opinion of experts, case studies, or standard-of-care." |
| Provide all other grades and definitions from the evidence grading system | See question above and next two questions below for more information. |
| Grade assigned to the **recommendation** with definition of the grade | The recommendations included have been assigned a Class I recommendation. Class I recommendations refer to "Conditions for which there is evidence and/or general agreement that a given procedure or treatment is beneficial, useful, and effective." |
| Provide all other grades and definitions from the recommendation grading system | ACCF/AHA guideline methodology categorizes indications as class I,II, or III on the basis of a multifactorial assessment of risk and expected efficacy viewed in the context of current knowledge and the relative strength of this knowledge. These classes summarize the recommendations for procedures or treatments as follows and noted in the table below:<br><br><u>Classification Types</u><br>Class I: Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective.<br><br>Class II: Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment.<br>• IIa: Weight of evidence/opinion is in favor of usefulness/efficacy<br>• IIb: Usefulness/efficacy is less well established by evidence/opinion.<br><br>Class III: Conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective e and in some cases may be harmful.<br>• No Benefit- Procedure/Test not helpful or Treatment w/o established proven benefit<br>• Harm- Procedure/Test leads to excess cost w/o benefit or is harmful, and or Treatment is harmful |

| | |
|---|---|
| | Additional detail regarding the classification of recommendation and level of evidence is provided in the following table.<br><br>Table 1:<br>  <br> |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of studies? | All but one of the recommendations for this process is rated as Level of Evidence A or B, meaning that the data was derived from one or more RCTs or meta-analyses. Additional information on the overall quality of evidence across the RCTs is not provided; although, three of the cited guidelines discuss the evidence supporting the use of beta blockers in this population, which is provided below.<br><br>2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes (p. e 159)<br><br>Beta blockers decrease heart rate, contractility, and BP, resulting in decreased $MVO_2$. Beta blockers without increased sympathomimetic activity should be administered orally in the absence of contraindications. Although early administration does not reduce short-term mortality (241,244), beta blockers decrease myocardial ischemia, reinfarction, and the frequency of complex ventricular dysrhythmias (240,245), and they increase long-term survival.<br><br>2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction (p. e104)<br><br>The benefit of beta blockers for secondary prevention has been established in numerous trials conducted in the prereperfusion era and appears to be greatest for patients with MI complicated by HF, LV dysfunction, or ventricular |

| | |
|---|---|
| | arrhythmias (418). The long-term duration of routine beta-blocker therapy after uncomplicated MI in patients without HF or hypertension has not been prospectively addressed. AHA/ ACCF secondary prevention guidelines recommend a 3-year treatment course in this patient subset (257).<br><br>2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease (p. e96)<br>Beta-receptor activation is associated with increases in heart rate, accelerated AV nodal conduction, and increased contractility, which contribute to increased myocardial oxygen demand. Decreases in the rate–BP product, AV nodal conduction, and myocardial contractility from beta blockers reduce myocardial oxygen demand, counteracting beta- receptor activity and contributing to a reduction in angina onset, with improvement in the ischemic threshold during exercise and in symptoms (764–769). These agents significantly reduce deaths and recurrent MIs in patients who have suffered a MI and are especially effective when a STEMI is complicated by persistent or recurrent ischemia or tachyarrhythmias early after the onset of infarction (757). However, no large trials have assessed effects of beta blockers on survival or coronary event rates in patients with SIHD. |
| Estimates of benefit and consistency across studies | Estimates of the benefit of beta blocker therapy across the body of evidence are not reported. |
| What harms were identified? | NA |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | Updated guidelines continue to support this measure. |

**1a.4 OTHER SOURCE OF EVIDENCE**

*If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.*

**1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure.** A list of references without a summary is not acceptable.

**1a.4.2 What process was used to identify the evidence?**

**1a.4.3. Provide the citation(s) for the evidence.**

---

**Measure Number** (*if previously endorsed*)**:** 0965

**Measure Title**: Patients with LVSD and an ICD implant who receive beta blocker therapy at discharge

**IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:** Patients with an ICD implant who receive ACE-I/ARB and beta blocker therapy at discharge

**Date of Submission**: Click here to enter a date

**1a.1.This is a measure of**: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome: Click here to name the health outcome

☐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*):

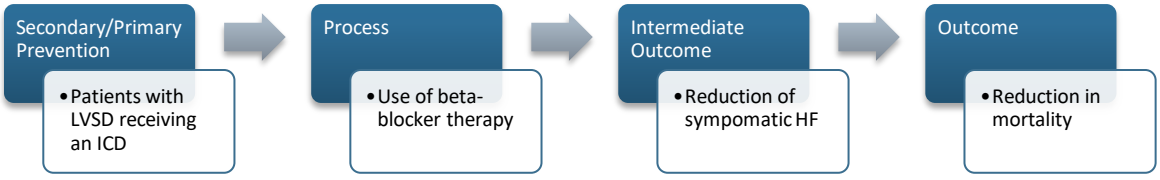
☒ Process: Beta-blocker therapy for patients with LVSD receiving an ICD

☐ 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.



Beta-blockers reduce morbidity, mortality, and hospitalizations for patients with heart failure and left ventricular systolic dysfunction.

**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.**

**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:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation,** | **Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJV, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WHW, Tsai EJ, Wilkoff BL. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62:e147–239.**<br><br>**http://content.onlinejacc.org/article.aspx?articleid=1695825** |

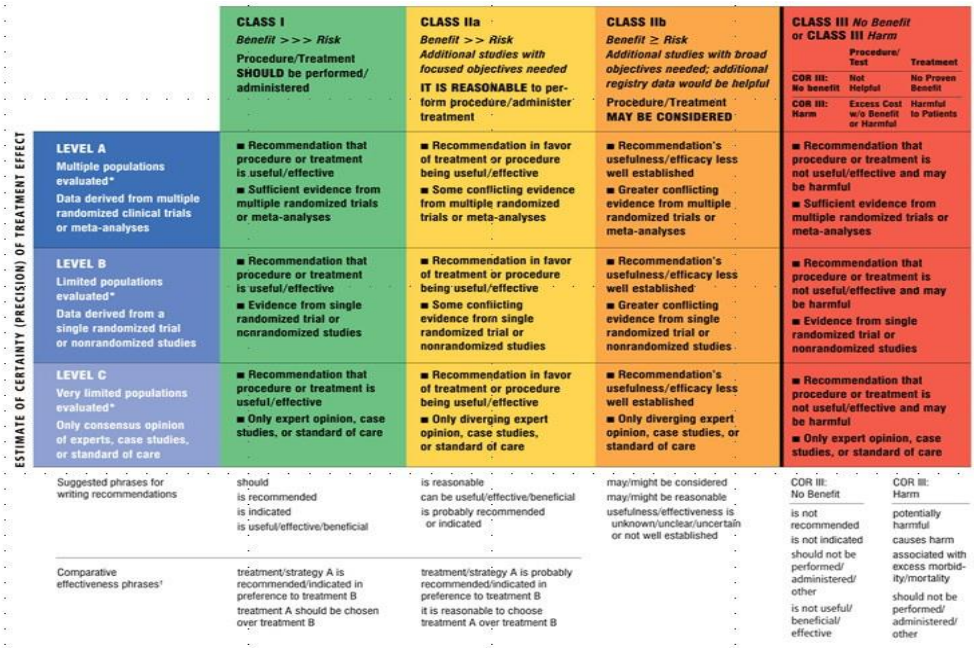
| | |
|---|---|
| • **including page number**<br>• **URL** | **Smith SC Jr., Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, Gibbons RJ, Grundy SM, Hiratzka LF, Jones DW, Lloyd-Jones DM, Minissian M, Mosca L, Peterson ED, Sacco RL, Spertus J, Stein JH, Taubert KA. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation. 2011: published online before print November 3, 2011, 10.1161/CIR.0b013e318235eb4d. http://content.onlinejacc.org/article.aspx?articleid=1147807**<br><br>**Al-Khatib SM, Stevenson WG, Ackerman MJ, Bryant WJ, Callans DJ, Curtis AB, Deal BJ, Dickfeld T, Field ME, Fonarow GC, Gillis AM, Granger CB, Hammill SC, Hlatky MA, Joglar JA, Kay GN, Matlock DD, Myerburg RJ, Page RL. 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. J Am Coll Cardiol 2018;72:e91–220. http://www.onlinejacc.org/content/accj/72/14/e91.full.pdf?_ga=2.238225088.1385433901.1570125164-1634948755.1534437338** |
| 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. | 2013 ACCF/AHA Guideline for the Management of Heart Failure (p. e169-170, 176, 195)<br>Stages of Heart Failure:<br>Stage A: At high risk for HF, but without structural heart disease or symptoms of Failure<br>Stage B: Structural heart disease, but without signs or symptoms of HF<br>Stage C: Structural heart disease with prior or current symptoms of HF<br>Stage D: Refractory HF requiring specialized interventions<br><br>p. e169<br>Stage B:<br>• In all patients with a recent or remote history of MI or ACS and reduced EF, evidence-based beta blockers should be used to reduce mortality. **Class I: Level of Evidence: B**<br>• Beta blockers should be used in all patients with a reduced EF to prevent symptomatic HF, even if they do not have a history of MI. **Class I: Level of Evidence: C**<br><br>p. e176:<br>Stage C:<br>• Use of 1 of the 3 beta blockers proven to reduce mortality (e.g., bisoprolol, carvedilol, and sustained-release metoprolol succinate) is recommended for all patients with current or prior symptoms of HFrEF, unless contraindicated, |

| | |
|---|---|
| | to reduce morbidity and mortality. **Class I: Level of Evidence: A**<br><br>AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update (p. 2435)<br><br>• Beta-blocker therapy should be used in all patients with left ventricular systolic dysfunction (ejection fraction <40%) with heart failure or prior myocardial infarction, unless contraindicated. (Use should be limited to carvedilol, metoprolol succinate, or bisoprolol, which have been shown to reduce mortality.) **Class I: Level of Evidence: A**<br><br>2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death (p. e116):<br><br>• In patients with HFrEF (LVEF ≤40%), treatment with a beta blocker, a mineralocorticoid receptor antagonist and either an angiotensin-converting enzyme inhibitor, an angiotensin-receptor blocker, or an angiotensin receptor-neprilysin inhibitor is recommended to reduce SCD and all-cause mortality. **Class I: Level of Evidence: A** |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | For guidelines released prior to 2015:<br>The weight of the evidence in support of most of the recommendations included are rated as Level A, Level B and Level C, as noted following each statement. Level A evidence refers to "Data derived from multiple randomized clinical trials or meta-analyses." The weight of the evidence in support of additional recommendations is rated as Level B and C. Level B evidence refers to "Data derived from a single randomized trial, or nonrandomized studies" while Level C evidence refers to "Only consensus opinion of experts, case studies, or standard-of-care."<br><br>For guidelines released from 2015 forward:<br>The weight of the evidence in support of most of the recommendations included are rated as Level A, Level B-R, Level B-NR, Level C-LD and Level C-EO, as noted following each statement. Level A evidence refers to high quality evidence from more than one randomized control trial (RCT), meta analyses of high-quality RCTs, and/or one or more RCTs corroborated by high-quality registry studies. Level B-R evidence refers to moderate-quality evidence from one or more RCTs and/or meta-analyses of moderate-quality RCTs and Level B-NR evidence includes moderate quality evidence from one or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies and/or meta-analyses of such studies. Level C-LD refers to randomized or nonrandomized observational or registry studies with limitation of design or execution, meta-analyses of such studies, and/or physiological or mechanistic studies in human subjects. Level C-EO refers to consensus of expert opinion based on clinical experience. |
| Provide all other grades | See question above and next two questions below for more information. |

| | |
|---|---|
| and definitions from the evidence grading system | |
| Grade assigned to the **recommendation** with definition of the grade | The recommendations included have been assigned a Class I recommendation.<br><br>For guidelines released prior to 2015:<br>Class I recommendations refer to "Conditions for which there is evidence and/or general agreement that a given procedure or treatment is beneficial, useful, and effective."<br><br>For guidelines released from 2015 forward:<br>Class I recommendations are "strong and indicate that the treatment, procedure, or intervention is useful and effective and should be performed or administered for most patients under most circumstances." |
| Provide all other grades and definitions from the recommendation grading system | For guidelines released prior to 2015:<br>ACCF/AHA guideline methodology categorizes indications as class I,II, or III on the basis of a multifactorial assessment of risk and expected efficacy viewed in the context of current knowledge and the relative strength of this knowledge. These classes summarize the recommendations for procedures or treatments as follows and noted in the table below:<br><br><u>Classification Types</u><br>Class I: Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective.<br><br>Class II: Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment.<br>• IIa: Weight of evidence/opinion is in favor of usefulness/efficacy<br>• IIb: Usefulness/efficacy is less well established by evidence/opinion.<br><br>Class III: Conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective e and in some cases may be harmful.<br>• No Benefit- Procedure/Test not helpful or Treatment w/o established proven benefit<br>• Harm- Procedure/Test leads to excess cost w/o benefit or is harmful, and or Treatment is harmful |

Additional detail regarding the classification of recommendation and level of evidence is provided in the following table.

Table 1:



| ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT | CLASS I<br>*Benefit >>> Risk*<br>Procedure/Treatment **SHOULD** be performed/ administered | CLASS IIa<br>*Benefit >> Risk*<br>*Additional studies with focused objectives needed*<br>**IT IS REASONABLE** to perform procedure/administer treatment | CLASS IIb<br>*Benefit ≥ Risk*<br>*Additional studies with broad objectives needed; additional registry data would be helpful*<br>Procedure/Treatment **MAY BE CONSIDERED** | CLASS III *No Benefit* or CLASS III *Harm*<br>COR III: No benefit — Procedure/Test: Not Helpful — Treatment: No Proven Benefit<br>COR III: Harm — Procedure/Test: Excess Cost w/o Benefit or Harmful — Treatment: Harmful to Patients |
|---|---|---|---|---|
| **LEVEL A**<br>Multiple populations evaluated*<br>Data derived from multiple randomized clinical trials or meta-analyses | ■ Recommendation that procedure or treatment is useful/effective<br>■ Sufficient evidence from multiple randomized trials or meta-analyses | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Some conflicting evidence from multiple randomized trials or meta-analyses | ■ Recommendation's usefulness/efficacy less well established<br>■ Greater conflicting evidence from multiple randomized trials or meta-analyses | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Sufficient evidence from multiple randomized trials or meta-analyses |
| **LEVEL B**<br>Limited populations evaluated*<br>Data derived from a single randomized trial or nonrandomized studies | ■ Recommendation that procedure or treatment is useful/effective<br>■ Evidence from single randomized trial or nonrandomized studies | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Some conflicting evidence from single randomized trial or nonrandomized studies | ■ Recommendation's usefulness/efficacy less well established<br>■ Greater conflicting evidence from single randomized trial or nonrandomized studies | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Evidence from single randomized trial or nonrandomized studies |
| **LEVEL C**<br>Very limited populations evaluated*<br>Only consensus opinion of experts, case studies, or standard of care | ■ Recommendation that procedure or treatment is useful/effective<br>■ Only expert opinion, case studies, or standard of care | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Only diverging expert opinion, case studies, or standard of care | ■ Recommendation's usefulness/efficacy less well established<br>■ Only diverging expert opinion, case studies, or standard of care | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Only expert opinion, case studies, or standard of care |
| Suggested phrases for writing recommendations | should<br>is recommended<br>is indicated<br>is useful/effective/beneficial | is reasonable<br>can be useful/effective/beneficial<br>is probably recommended or indicated | may/might be considered<br>may/might be reasonable<br>usefulness/effectiveness is unknown/unclear/uncertain or not well established | COR III: No Benefit<br>is not recommended<br>is not indicated<br>should not be performed/ administered/ other<br>is not useful/ beneficial/ effective<br><br>COR III: Harm<br>potentially harmful<br>causes harm<br>associated with excess morbidity/mortality<br>should not be performed/ administered/ other |
| Comparative effectiveness phrases† | treatment/strategy A is recommended/indicated in preference to treatment B<br>treatment A should be chosen over treatment B | treatment/strategy A is probably recommended/indicated in preference to treatment B<br>it is reasonable to choose treatment A over treatment B | | |

For guidelines released from 2015 forward:

In 2015, the ACC and AHA updated Classes of Recommendation (COR) and Levels of Evidence (LOE) in an effort to align patient care with scientific evidence.

The COR reflects the magnitude of benefit over risk and corresponds to the strength of the recommendation. Class I recommendations are strong and indicate that the treatment, procedure, or intervention is useful and effective and should be performed or administered for most patients under most circumstances. Class II recommendations are weaker, denoting a lower degree of benefit in proportion to risk. Benefit is generally greater for Class IIa (moderate) recommendations and smaller for Class IIb (weak) recommendations, for which benefit only marginally exceeds risk. A COR of IIb suggests that implementation should be selective and based on careful consideration of individual patient factors and, for invasive procedures, available expertise. Class III is assigned when actions are specifically not recommended, either because studies have found no evidence of benefit or because the intervention causes harm.

Additional detail regarding the COR and LOE is provided in the following table.

| | |
|---|---|
| | Table 2:<br>  <br> |
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of | All but one of the recommendations for this process is rated as Level of Evidence A or B, meaning that the data was derived from one or more RCTs or meta-analyses. Additional information on the overall quality of evidence across the RCTs is not provided; although, two of the cited guidelines discuss the evidence supporting the use of beta blockers in this population, which is provided below.<br><br>2013 ACCF/AHA Guideline for the Management of Heart Failure (p. e169-170, 176, 195)<br>The body of evidence supporting the recommendations on beta-blocker therapy for patients with LVSD in this guideline includes 7 randomized controlled trials.<br><br>p. e170:<br>CAD is a major risk factor for the development of HF and a key target for |

| | |
|---|---|
| studies? | prevention of HF. The 5-year risk of developing HF after acute MI is 7% and 12% for men and women, respectively; for men and women between the ages of 40 and 69 and those >70 years of age, the risk is 22% and 25%, respectively (51). Current evidence supports the use of ACE inhibitors and (to a lower level of evidence) beta-blocker therapy to impede maladaptive LV remodeling in patients with stage B HF and low LVEF to improve mortality and morbidity (344).<br><br>In 1 study, losartan reduced adverse outcomes in a population with hypertension (357), and in another study of patients post-MI with low LVEF, valsartan was equivalent to captopril (345). Data with beta blockers are less convincing in a population with known CAD, although in 1 trial (346) carvedilol therapy in patients with stage B and low LVEF was associated with a 31% relative risk reduction in adverse long-term outcomes. In patients with previously established structural heart disease, the administration of agents known to have negative inotropic properties such as non-dihydropyridine calcium channel blockers and certain antiarrhythmics should be avoided.<br><br>p. e176:<br><br>Three beta blockers have been shown to be effective in reducing the risk of death in patients with chronic HFrEF: bisoprolol and sustained-release metoprolol (succinate), which selectively block beta-1–receptors; and carvedilol, which blocks alpha-1–, beta-1–, and beta-2–receptors. Positive findings with these 3 agents, however, should not be considered a beta-blocker class effect. Bucindolol lacked uniform effectiveness across different populations, and short-acting metoprolol tartrate was less effective in HF clinical trials. Beta-1 selective blocker nebivolol demonstrated a modest reduction in the primary endpoint of all-cause mortality or cardiovascular hospitalization but did not affect mortality alone in an elderly population that included patients with HFpEF (472).<br><br>2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death (p. e116):<br><br>RCTs in patients with HFrEF have consistently demonstrated that chronic therapy with beta blockers reduces all-cause mortality, VA, and SCD (S5.2-2,S5.2-4, S5.2-5,S5.2-9). Three beta blockers (i.e., bisoprolol, carvedilol, sustained-release metoprolol succinate) have been proven to reduce mortality in patients with current or prior symptoms of HFrEF without beta-blocker contraindications. |
| Estimates of benefit and consistency across studies | Estimates of the benefit of beta blocker therapy across the body of evidence are not reported. |
| What harms were identified? | NA |

| | |
|---|---|
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | Updated guidelines continue to support this measure. |

**1a.4 OTHER SOURCE OF EVIDENCE**

*If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.*

**1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure.** A list of references without a summary is not acceptable.

**1a.4.2 What process was used to identify the evidence?**

**1a.4.3. Provide the citation(s) for the evidence.**

**Component # 2 ACE-I/ARB Therapy**

**Measure Number** (*if previously endorsed*)**:** 0965

**Measure Title**: Patients with LVSD and an ICD implant who receive ACE-I/ARB therapy at discharge

**IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:** Patients with an ICD implant who receive ACE-I/ARB and beta blocker therapy at discharge

**Date of Submission**: Click here to enter a date

**1a.1.This is a measure of**: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome: Click here to name the health outcome

☐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*):

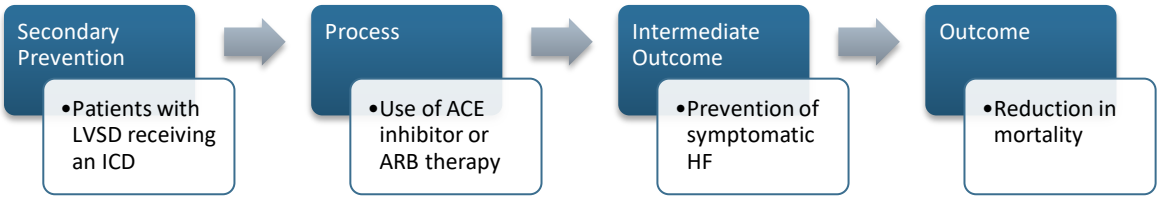
☒ Process: ACE/ARB therapy for patients with LVSD receiving an ICD

☐ 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.




Angiotensin-converting enzyme (ACE) inhibitors and angiotensin-receptor antagonists/blockers (ARBs) reduce morbidity, mortality, and hospitalizations for patients with heart failure and left ventricular systolic dysfunction.

**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.**

**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:**<br>• **Title**<br>• **Author**<br>• **Date**<br>• **Citation, including page number**<br>• **URL** | **Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJV, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WHW, Tsai EJ, Wilkoff BL. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62:e147–239.**<br><br>**http://content.onlinejacc.org/article.aspx?articleid=1695825**<br><br>**Smith SC Jr., Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, Gibbons RJ, Grundy SM, Hiratzka LF, Jones DW, Lloyd-Jones DM, Minissian M, Mosca L, Peterson ED, Sacco RL, Spertus J, Stein JH, Taubert KA. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation. 2011: published online before print November 3, 2011, 10.1161/CIR.0b013e318235eb4d.**<br>**http://content.onlinejacc.org/article.aspx?articleid=1147807**<br><br>**Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Colvin MM, Drazner MH, Filippatos GS, Fonarow GC, Givertz MM, Hollenberg SM, Lindenfeld J, Masoudi FA, McBride PE, Peterson PN, Stevenson LW, Westlake C. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. J Am Coll Cardiol. 2017;70:776–803.**<br>**http://www.onlinejacc.org/content/accj/72/14/e91.full.pdf?_ga=2.238225088.1385433901.1570125164-1634948755.1534437338**<br><br>**Al-Khatib SM, Stevenson WG, Ackerman MJ, Bryant WJ, Callans DJ, Curtis AB, Deal BJ, Dickfeld T, Field ME, Fonarow GC, Gillis AM, Granger CB, Hammill SC, Hlatky MA, Joglar JA, Kay GN, Matlock DD, Myerburg RJ, Page RL. 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Clinical Practice Guidelines** |

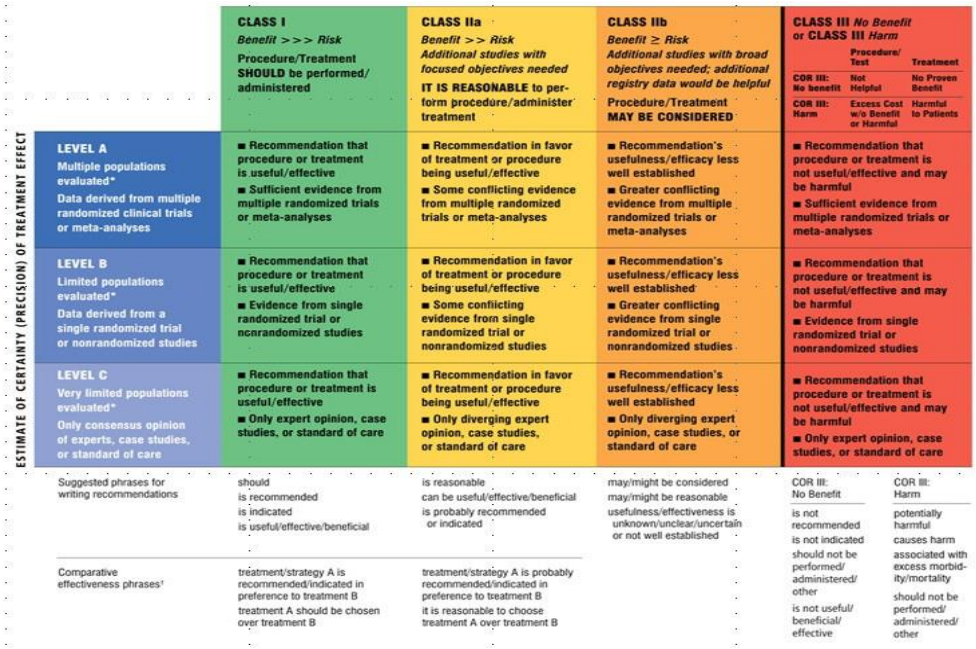
| | |
|---|---|
| | **and the Heart Rhythm Society. J Am Coll Cardiol 2018;72:e91–220.** **http://www.onlinejacc.org/content/accj/72/14/e91.full.pdf?_ga=2.238225088.1385433901.1570125164-1634948755.1534437338** |
| Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR. | 2013 ACCF/AHA Guideline for the Management of Heart Failure (e169-170, 174-175, 195)<br>Stages of Heart Failure:<br>Stage A: At high risk for HF, but without structural heart disease or symptoms of Failure<br>Stage B: Structural heart disease, but without signs or symptoms of HF<br>Stage C: Structural heart disease with prior or current symptoms of HF<br>Stage D: Refractory HF requiring specialized interventions<br><br>Stage B recommendations (e169-170):<br>• In all patients with a recent or remote history of MI or ACS and reduced EF, ACE inhibitors should be used to prevent symptomatic HF and reduce mortality. In patients intolerant of ACE inhibitors, ARBs are appropriate unless contraindicated. **Class I; Level of Evidence: A**<br>• ACE inhibitors should be used in all patients with a reduced EF to prevent symptomatic HF, even if they do not have a history of MI. **Class I; Level of Evidence: A**<br><br>AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update (p. 2435)<br>• ACE inhibitors should be started and continued indefinitely in all patients with left ventricular ejection fraction <40% and in those with hypertension, diabetes, or chronic kidney disease, unless contraindicated. **Class I; Level of Evidence: A**<br>• The use of ARBs is recommended in patients who have heart failure or who have had a myocardial infarction with left ventricular ejection fraction 40% and who are ACE-inhibitor intolerant. **Class I; Level of Evidence: A**<br><br>2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure (p. e784):<br>• The clinical strategy of inhibition of the renin- angiotensin system with ACE inhibitors (128–133), OR ARBs (134–137), OR ARNI (138) in conjunction with evidence-based beta blockers (9,139,140), and aldosterone antagonists in selected patients (141,142), is recommended for patients with chronic HFrEF to reduce morbidity and mortality. **Class I; ACE inhibitor: Level of Evidence: A, ARBs: Level of Evidence: A or ARNI: Level of Evidence: B-R** |

| | |
|---|---|
| | • The use of ACE inhibitors is beneficial for patients with prior or current symptoms of chronic HFrEF to reduce morbidity and mortality. **Class I; Level of Evidence: A**<br>• The use of ARBs to reduce morbidity and mortality is recommended in patients with prior or current symptoms of chronic HFrEF who are intolerant to ACE inhibitors because of cough or angioedema. **Class I; Level of Evidence: A**<br>• In patients with chronic symptomatic HFrEF NYHA class II or III who tolerate an ACE inhibitor or ARB, replacement by an ARNI is recommended to further reduce morbidity and mortality (138). **Class I; Level of Evidence: B-R**<br><br>2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death (p. e116):<br><br>• In patients with HFrEF (LVEF ≤40%), treatment with a beta blocker, a mineralocorticoid receptor antagonist and either an angiotensin-converting enzyme inhibitor, an angiotensin-receptor blocker, or an angiotensin receptor-neprilysin inhibitor is recommended to reduce SCD and all-cause mortality. **Class I: Level of Evidence: A** |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade | The weight of the evidence in support of most of the recommendations included are rated as Level A, Level B and Level C, as noted following each statement. Level A evidence refers to "Data derived from multiple randomized clinical trials or meta-analyses." The weight of the evidence in support of additional recommendations is rated as Level B and C. Level B evidence refers to "Data derived from a single randomized trial, or nonrandomized studies" while Level C evidence refers to "Only consensus opinion of experts, case studies, or standard-of-care."<br><br>For guidelines released from 2015 forward:<br>The weight of the evidence in support of most of the recommendations included are rated as Level A, Level B-R, Level B-NR, Level C-LD and Level C-EO, as noted following each statement. Level A evidence refers to high quality evidence from more than one randomized control trial (RCT), meta analyses of high-quality RCTs, and/or one or more RCTs corroborated by high-quality registry studies. Level B-R evidence refers to moderate-quality evidence from one or more RCTs and/or meta-analyses of moderate-quality RCTs and Level B-NR evidence includes moderate quality evidence from one or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies and/or meta-analyses of such studies. Level C-LD refers to randomized or nonrandomized observational or registry studies with limitation of design or execution, meta-analyses of such studies, and/or physiological or mechanistic studies in human subjects. Level C-EO refers to consensus of expert opinion based on clinical experience. |
| Provide all other grades | See question above and next two questions below for more information. |

| | |
|---|---|
| and definitions from the evidence grading system | |
| Grade assigned to the **recommendation** with definition of the grade | The recommendations included have been assigned a Class I recommendation.<br><br>For guidelines released prior to 2015:<br>Class I recommendations refer to "Conditions for which there is evidence and/or general agreement that a given procedure or treatment is beneficial, useful, and effective."<br><br>For guidelines released from 2015 forward:<br>Class I recommendations are "strong and indicate that the treatment, procedure, or intervention is useful and effective and should be performed or administered for most patients under most circumstances." |
| Provide all other grades and definitions from the recommendation grading system | For guidelines released prior to 2015:<br>ACCF/AHA guideline methodology categorizes indications as class I,II, or III on the basis of a multifactorial assessment of risk and expected efficacy viewed in the context of current knowledge and the relative strength of this knowledge. These classes summarize the recommendations for procedures or treatments as follows and noted in the table below:<br><br><u>Classification Types</u><br>Class I: Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective.<br><br>Class II: Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment.<br>• IIa: Weight of evidence/opinion is in favor of usefulness/efficacy<br>• IIb: Usefulness/efficacy is less well established by evidence/opinion.<br><br>Class III: Conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective e and in some cases may be harmful.<br>• No Benefit- Procedure/Test not helpful or Treatment w/o established proven benefit<br>• Harm- Procedure/Test leads to excess cost w/o benefit or is harmful, and or Treatment is harmful |

Additional detail regarding the classification of recommendation and level of evidence is provided in the following table.

Table 1:



| ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT | CLASS I<br>*Benefit >>> Risk*<br>Procedure/Treatment **SHOULD** be performed/ administered | CLASS IIa<br>*Benefit >> Risk*<br>*Additional studies with focused objectives needed*<br>**IT IS REASONABLE** to perform procedure/administer treatment | CLASS IIb<br>*Benefit ≥ Risk*<br>*Additional studies with broad objectives needed; additional registry data would be helpful*<br>Procedure/Treatment **MAY BE CONSIDERED** | CLASS III *No Benefit* or CLASS III *Harm*<br>COR III: No benefit — Procedure/Test: Not Helpful — Treatment: No Proven Benefit<br>COR III: Harm — Procedure/Test: Excess Cost w/o Benefit or Harmful — Treatment: Harmful to Patients |
|---|---|---|---|---|
| **LEVEL A**<br>Multiple populations evaluated*<br>Data derived from multiple randomized clinical trials or meta-analyses | ■ Recommendation that procedure or treatment is useful/effective<br>■ Sufficient evidence from multiple randomized trials or meta-analyses | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Some conflicting evidence from multiple randomized trials or meta-analyses | ■ Recommendation's usefulness/efficacy less well established<br>■ Greater conflicting evidence from multiple randomized trials or meta-analyses | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Sufficient evidence from multiple randomized trials or meta-analyses |
| **LEVEL B**<br>Limited populations evaluated*<br>Data derived from a single randomized trial or nonrandomized studies | ■ Recommendation that procedure or treatment is useful/effective<br>■ Evidence from single randomized trial or nonrandomized studies | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Some conflicting evidence from single randomized trial or nonrandomized studies | ■ Recommendation's usefulness/efficacy less well established<br>■ Greater conflicting evidence from single randomized trial or nonrandomized studies | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Evidence from single randomized trial or nonrandomized studies |
| **LEVEL C**<br>Very limited populations evaluated*<br>Only consensus opinion of experts, case studies, or standard of care | ■ Recommendation that procedure or treatment is useful/effective<br>■ Only expert opinion, case studies, or standard of care | ■ Recommendation in favor of treatment or procedure being useful/effective<br>■ Only diverging expert opinion, case studies, or standard of care | ■ Recommendation's usefulness/efficacy less well established<br>■ Only diverging expert opinion, case studies, or standard of care | ■ Recommendation that procedure or treatment is not useful/effective and may be harmful<br>■ Only expert opinion, case studies, or standard of care |
| Suggested phrases for writing recommendations | should<br>is recommended<br>is indicated<br>is useful/effective/beneficial | is reasonable<br>can be useful/effective/beneficial<br>is probably recommended or indicated | may/might be considered<br>may/might be reasonable<br>usefulness/effectiveness is unknown/unclear/uncertain or not well established | COR III: No Benefit<br>is not recommended<br>is not indicated<br>should not be performed/ administered/ other<br>is not useful/ beneficial/ effective<br><br>COR III: Harm<br>potentially harmful<br>causes harm<br>associated with excess morbidity/mortality<br>should not be performed/ administered/ other |
| Comparative effectiveness phrases† | treatment/strategy A is recommended/indicated in preference to treatment B<br>treatment A should be chosen over treatment B | treatment/strategy A is probably recommended/indicated in preference to treatment B<br>it is reasonable to choose treatment A over treatment B | | |

For guidelines released from 2015 forward:

In 2015, the ACC and AHA updated Classes of Recommendation (COR) and Levels of Evidence (LOE) in an effort to align patient care with scientific evidence.

The COR reflects the magnitude of benefit over risk and corresponds to the strength of the recommendation. Class I recommendations are strong and indicate that the treatment, procedure, or intervention is useful and effective and should be performed or administered for most patients under most circumstances. Class II recommendations are weaker, denoting a lower degree of benefit in proportion to risk. Benefit is generally greater for Class IIa (moderate) recommendations and smaller for Class IIb (weak) recommendations, for which benefit only marginally exceeds risk. A COR of IIb suggests that implementation should be selective and based on careful consideration of individual patient factors and, for invasive procedures, available expertise. Class III is assigned when actions are specifically not recommended, either because studies have found no evidence of benefit or because the intervention causes harm.

Additional detail regarding the COR and LOE is provided in the following table.

Table 2:

**CLASS (STRENGTH) OF RECOMMENDATION**

**CLASS I (STRONG)** — Benefit >>> Risk

Suggested phrases for writing recommendations:
- Is recommended
- Is indicated/useful/effective/beneficial
- Should be performed/administered/other
- Comparative-Effectiveness Phrases†:
  - Treatment/strategy A is recommended/indicated in preference to treatment B
  - Treatment A should be chosen over treatment B

**CLASS IIa (MODERATE)** — Benefit >> Risk

Suggested phrases for writing recommendations:
- Is reasonable
- Can be useful/effective/beneficial
- Comparative-Effectiveness Phrases†:
  - Treatment/strategy A is probably recommended/indicated in preference to treatment B
  - It is reasonable to choose treatment A over treatment B

**CLASS IIb (WEAK)** — Benefit ≥ Risk

Suggested phrases for writing recommendations:
- May/might be reasonable
- May/might be considered
- Usefulness/effectiveness is unknown/unclear/uncertain or not well established

**CLASS III: No Benefit (MODERATE)** — Benefit = Risk
*(Generally, LOE A or B use only)*

Suggested phrases for writing recommendations:
- Is not recommended
- Is not indicated/useful/effective/beneficial
- Should not be performed/administered/other

**CLASS III: Harm (STRONG)** — Risk > Benefit

Suggested phrases for writing recommendations:
- Potentially harmful
- Causes harm
- Associated with excess morbidity/mortality
- Should not be performed/administered/other

**LEVEL (QUALITY) OF EVIDENCE‡**

**LEVEL A**
- High-quality evidence‡ from more than 1 RCT
- Meta-analyses of high-quality RCTs
- One or more RCTs corroborated by high-quality registry studies

**LEVEL B-R** (Randomized)
- Moderate-quality evidence‡ from 1 or more RCTs
- Meta-analyses of moderate-quality RCTs

**LEVEL B-NR** (Nonrandomized)
- Moderate-quality evidence‡ from 1 or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies
- Meta-analyses of such studies

**LEVEL C-LD** (Limited Data)
- Randomized or nonrandomized observational or registry studies with limitations of design or execution
- Meta-analyses of such studies
- Physiological or mechanistic studies in human subjects

**LEVEL C-EO** (Expert Opinion)

Consensus of expert opinion based on clinical experience

COR and LOE are determined independently (any COR may be paired with any LOE).

A recommendation with LOE C does not imply that the recommendation is weak. Many important clinical questions addressed in guidelines do not lend themselves to clinical trials. Although RCTs are unavailable, there may be a very clear clinical consensus that a particular test or therapy is useful or effective.

* The outcome or result of the intervention should be specified (an improved clinical outcome or increased diagnostic accuracy or incremental prognostic information).

† For comparative-effectiveness recommendations (COR I and IIa; LOE A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated.

‡ The method of assessing quality is evolving, including the application of standardized, widely used, and preferably validated evidence grading tools; and for systematic review the incorporation of an Evidence Review Committee.

COR indicates Class of Recommendation; EO, expert opinion; LD, limited data; LOE, Level of Evidence; NR, nonrandomized; R, randomized; and RCT, randomized controlled trial.

| | |
|---|---|
| Body of evidence:<br>• Quantity – how many studies?<br>• Quality – what type of | All of the recommendations for this process are rated as Level of Evidence A or B, meaning that the data was derived from one or more RCTs or meta-analyses. Additional information on the overall quality of evidence across the RCTs is not provided; although, three of the cited guidelines discuss the evidence supporting the use of ACE inhibitors or ARBs in this population, which is provided below.<br><br>2013 ACCF/AHA Guideline for the Management of Heart Failure (p. e170)<br>The body of evidence supporting the recommendations on ACE/ARB therapy in this guideline includes 15 randomized controlled trials.<br><br>CAD is a major risk factor for the development of HF and a key target for prevention of HF. The 5-year risk of developing HF after acute MI is 7% and 12% for men and women, respectively; for men and women between the ages of 40 and 69 and those >70 years of age, the risk is 22% and 25%, respectively (51). |

| | |
|---|---|
| studies? | Current evidence supports the use of ACE inhibitors and (to a lower level of evidence) beta-blocker therapy to impede maladaptive LV remodeling in patients with stage B HF and low LVEF to improve mortality and morbidity (344). At 3-year follow-up, those patients treated with ACE inhibitors demonstrated combined endpoints of reduced hospitalization or death, a benefit that extended up to a 12-year follow-up (65). ARBs are reasonable alternatives to ACE inhibitors.<br><br>2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure (p. e784):<br><br>Angiotensin-converting enzyme (ACE) inhibitors reduce morbidity and mortality in heart failure with reduced ejection fraction (HFrEF). Randomized controlled trials (RCTs) clearly establish the benefits of ACE inhibition in patients with mild, moderate, or severe symptoms of HF and in patients with or without coronary artery disease (128–133). ACE inhibitors can produce angioedema and should be given with caution to patients with low systemic blood pressures, renal insufficiency, or elevated serum potassium. ACE inhibitors also inhibit kininase and increase levels of bradykinin, which can induce cough but also may contribute to their beneficial effect through vasodilation.<br><br>Angiotensin receptor blockers (ARBs) were developed with the rationale that angiotensin II production continues in the presence of ACE inhibition, driven through alternative enzyme pathways. ARBs do not inhibit kininase and are associated with a much lower incidence of cough and angioedema than ACE inhibitors; but like ACE inhibitors, ARBs should be given with caution to patients with low systemic blood pressure, renal insufficiency, or elevated serum potassium. Long-term therapy with ARBs produces hemodynamic, neurohormonal, and clinical effects consistent with those expected after interference with the renin-angiotensin system and have been shown in RCTs (134–137) to reduce morbidity and mortality, especially in ACE inhibitor–intolerant patients.<br><br>Benefits of ACE inhibitors with regard to decreasing HF progression, hospitalizations, and mortality rate have been shown consistently for patients across the clinical spectrum, from asymptomatic to severely symptomatic HF. Similar benefits have been shown for ARBs in populations with mild-to-moderate HF who are unable to tolerate ACE inhibitors. In patients with mild-to-moderate HF (characterized by either 1) mildly elevated natriuretic peptide levels, BNP [B-type natriuretic peptide]≥150 pg/mL or NT-proBNP [N-terminal pro-B-type natriuretic peptide] ≥600 pg/mL; or 2) BNP≥100 pg/mL or NT-proBNP≥400 pg/mL with a prior hospitalization in the preceding 12 months) who were able to tolerate both a target dose of enalapril (10 mg twice daily) and then subsequently an ARNI (valsartan/sacubitril; 200 mg twice daily, with the ARB component equivalent to valsartan160 mg), hospitalizations and mortality were significantly decreased with the valsartan/sacubitril compound compared with enalapril. The target dose of the ACE inhibitor was consistent with that known to improve outcomes in previous landmark clinical trials(129). This ARNI has been approved for patients with symptomatic HFrEF and is intended to be substituted for ACE inhibitors or |

| | |
|---|---|
| | ARBs. HF effects and potential off-target effects maybe complex with inhibition of the neprilysin enzyme, which has multiple biological targets. Use of an ARNI is associated with hypotension and a low-frequency incidence of angioedema.<br><br>2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death (p. e116):<br><br>Angiotensin-converting enzyme inhibition also reduces mortality and SCD (S5.2-3). Angiotensin-receptor blockers added to angiotensin-converting enzyme inhibitor showed additional benefit to angiotensin-converting enzyme inhibitors in some (S5.2-10) but not other RCTs (S5.2-8,S5.2-11). Therapy with the mineralocorticoid-receptor antagonists, spironolactone and eplerenone, have also demonstrated reductions in both all-cause mortality and SCD (S5.2-6,S5.2-12,S5.2-13). Recent studies of the angiotensin receptor-neprilysin inhibitor (sacubitril/valsartan) versus angiotensin-converting enzyme inhibitor demonstrated a reduction in SCD and cardiac mortality (S5.2-14). |
| Estimates of benefit and consistency across studies | Estimates of the benefit of ACE/ARB therapy across the body of evidence are not reported. |
| What harms were identified? | NA |
| Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR? | Updated guidelines continue to support this measure. |

---

**1a.4 OTHER SOURCE OF EVIDENCE**

*If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.*

**1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure.** A list of references without a summary is not acceptable.

**1a.4.2 What process was used to identify the evidence?**

**1a.4.3. Provide the citation(s) for the evidence.**

## 1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

**1b.1. Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

*If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.*

This measure is intended to assess the extent to which eligible patients receive evidence-based medications that are indicated at hospital discharge following ICD placement. This measure focuses on processes of care that are supported by guidelines for optimal care for patients undergoing ICD placement.

**1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Performance scores from the National Cardiovascular Data Registry's ICD Registry, a national quality improvement registry are provided below. Performance scores from 2017 and 2018 are provided from the US hospitals participating in the registry.

Measurement Year: 2017

Number of facilities: 1,674

Mean: 82.72%

Std Dev: 18.00%

100% Max: 100.00%

99%: 100.00%

95%: 100.00%

90%: 100.00%

75% Q3: 96.84%

50% Median: 87.56%

25% Q1: 74.29%

10%: 60.00%

5%: 48.39%

1%: 0.00%

0% Min: 0.00%

Measurement Year: 2018

Number of facilities: 1,574

Mean: 82.54%

Std Dev: 19.80%

100% Max: 100.00%

99%: 100.00%

95%: 100.00%

90%: 100.00%

75% Q3: 97.41%

50% Median: 87.95%

25% Q1: 75.00%

10%: 59.80%

5%: 46.67%

1%: 0.00%

0% Min: 0.00%

**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.*

Performance scores from the National Cardiovascular Data Registry's ICD Registry, a national quality improvement registry are provided below. Performance scores from 2017 and 2018 stratified by gender, age, race/ethnicity, and dual eligibility are provided from the US hospitals participating in the registry.

Measurement Year: 2017

Number of facilities: 1,674

Gender - Male:

Mean: 82.88%

Std Dev: 18.50%

100% Max: 100.00%

99%: 100.00%

95%: 100.00%
90%: 100.00%
75% Q3: 97.30%
50% Median: 88.00%
25% Q1: 75.00%
10%: 60.24%
5%: 50.00%
1%: 0.00%
0% Min: 0.00%
Gender - Female:
Mean: 82.86%
Std Dev: 21.45%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 90.00%
25% Q1: 75.00%
10%: 50.00%
5%: 40.00%
1%: 0.00%
0% Min: 0.00%
Age <65:
Mean: 84.18%
Std Dev: 20.33%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 91.04%
25% Q1: 76.92%
10%: 58.06%
5%: 50.00%
1%: 0.00%

0% Min: 0.00%
Age =>65:
Mean: 81.99%
Std Dev: 19.25%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 97.50%
50% Median: 86.96%
25% Q1: 72.73%
10%: 58.82%
5%: 44.44%
1%: 0.00%
0% Min: 0.00%
Hispanic:
Mean: 84.06%
Std Dev: 26.94%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 100.00%
25% Q1: 76.47%
10%: 50.00%
5%: 0.00%
1%: 0.00%
0% Min: 0.00%
White non-Hispanic:
Mean: 82.80%
Std Dev: 18.37%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%

75% Q3: 97.37%
50% Median: 87.87%
25% Q1: 73.68%
10%: 60.00%
5%: 50.00%
1%: 0.00%
0% Min: 0.00%
Black non-Hispanic:
Mean: 84.41%
Std Dev: 23.59%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 96.67%
25% Q1: 76.46%
10%: 50.00%
5%: 33.33%
1%: 0.00%
0% Min: 0.00%
Other:
Mean: 86.46%
Std Dev: 27.07%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 100.00%
25% Q1: 85.71%
10%: 50.00%
5%: 0.00%
1%: 0.00%
0% Min: 0.00%
Non-White:

Mean: 82.10%
Std Dev: 18.99%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 96.55%
50% Median: 87.50%
25% Q1: 73.53%
10%: 57.14%
5%: 46.54%
1%: 0.00%
0% Min: 0.00%
Dual Eligibility:
Mean: 82.68%
Std Dev: 16.24%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 95.74%
50% Median: 86.67%
25% Q1: 73.77%
10%: 61.54%
5%: 50.00%
1%: 30.00%
0% Min: 0.00%
Measurement Year: 2018
Number of facilities: 1,574
Gender - Male:
Mean: 82.84%
Std Dev: 20.00%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%

75% Q3: 98.04%
50% Median: 88.89%
25% Q1: 75.00%
10%: 58.14%
5%: 47.62%
1%: 0.00%
0% Min: 0.00%
Gender - Female:
Mean: 82.43%
Std Dev: 22.90%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 91.43%
25% Q1: 75.00%
10%: 52.94%
5%: 33.33%
1%: 0.00%
0% Min: 0.00%
Age <65:
Mean: 85.22%
Std Dev: 20.87%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 92.98%
25% Q1: 78.00%
10%: 60.00%
5%: 50.00%
1%: 0.00%
0% Min: 0.00%
Age =>65:

Mean: 81.78%
Std Dev: 20.91%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 98.33%
50% Median: 88.00%
25% Q1: 72.92%
10%: 55.17%
5%: 42.86%
1%: 0.00%
0% Min: 0.00%
Hispanic:
Mean: 84.76%
Std Dev: 26.16%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 100.00%
25% Q1: 77.42%
10%: 50.00%
5%: 0.00%
1%: 0.00%
0% Min: 0.00%
White non-Hispanic:
Mean: 82.68%
Std Dev: 20.47%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 98.53%
50% Median: 88.75%

25% Q1: 74.83%
10%: 58.33%
5%: 43.75%
1%: 0.00%
0% Min: 0.00%
Black non-Hispanic:
Mean: 84.84%
Std Dev: 24.60%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 100.00%
25% Q1: 78.42%
10%: 50.00%
5%: 33.33%
1%: 0.00%
0% Min: 0.00%
Other:
Mean: 86.02%
Std Dev: 28.28%
100% Max: 100.00%
99%: 100.00%
95%: 100.00%
90%: 100.00%
75% Q3: 100.00%
50% Median: 100.00%
25% Q1: 85.71%
10%: 50.00%
5%: 0.00%
1%: 0.00%
0% Min: 0.00%
Non-White:
Mean: 82.48%
Std Dev: 20.07%

100% Max: 100.00%

99%: 100.00%

95%: 100.00%

90%: 100.00%

75% Q3: 97.22%

50% Median: 88.46%

25% Q1: 75.00%

10%: 55.56%

5%: 42.11%

1%: 0.00%

0% Min: 0.00%

Dual Eligibility:

Mean: 83.04%

Std Dev: 15.94%

100% Max: 100.00%

99%: 100.00%

95%: 100.00%

90%: 100.00%

75% Q3: 96.15%

50% Median: 86.41%

25% Q1: 75.00%

10%: 61.54%

5%: 53.13%

1%: 33.33%

0% Min: 0.00%

**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

## 1c. Composite Quality Construct and Rationale

**1c.1. A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.

- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
    - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient);

**1c.1.** Please identify the composite measure construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)

**1c.2. Describe the quality construct, including:**

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

This measure focuses on processes of care that recommended for optimal care for patients following ICD/CRT-D implantation. Each component of the measure has been shown in randomized clinical trials to impact clinical outcomes and represents a Class 1 guideline indication for the care of patients with left ventricular systolic dysfunction or prior myocardial infarction. Combining the individual process measures into a single composite provides patients, physicians, and hospitals with a perspective of the overall quality of medical therapy provided to patients undergoing ICD/CRT-D implantation. Hospitals with a gap in performance can investigate the individual components of the measure to identify specific opportunities for improvement. The content validity of this measure has been achieved by virtue of their consistency with strong guideline recommendations and the expertise of the individuals who developed this measure.

In addition, we conducted empiric analyses examining the association between performance on the composite measure and clinical outcomes including readmission and mortality at 6 months following device implantation (see testing supplement for detailed results). We found that patients who were discharged on appropriate medical therapy were less likely to experience adverse outcomes compared with patients who were not discharged on appropriate medical therapy. Furthermore, fewer patients treated at high performing hospitals as determined by this composite experienced adverse outcomes compared with those treated at low performing hospitals.

In addition, we conducted empiric analyses examining the association between performance on the composite measure and clinical outcomes including readmission and mortality at 6 months following device implantation (see testing supplement for detailed results). We found that patients who were discharged on appropriate medical therapy were less likely to experience adverse outcomes compared with patients who were not discharged on appropriate medical therapy. Furthermore, fewer patients treated at high performing hospitals as determined by this composite experienced adverse outcomes compared with those treated at low performing hospitals.

**1c.3. Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.**

This measure is intended to assess the extent to which eligible patients receive evidence-based medications that are indicated at hospital discharge following ICD/CRT-D implantation.

Composite performance measures have a variety of uses.

Data reduction. A large and growing array of individual indicators makes it possible for users to become overloaded with data. A composite measure reduces the information burden by distilling the available indicators into a simple summary.

Scope expansion. The information in a composite measure is highly condensed, making it feasible to track a broader range of metrics than would be possible otherwise. Composite measures have been described as a tool for making provider assessments more comprehensive.

Provider performance valuation. Performance indicators are used for various decisions about providers, including the allocation of pay-for-performance incentives, designation of preferred provider status, and assignment of letter grades and star rating categories. If a decision is to be based on multiple indicators instead of a single indicator, a method of translating several variables into a single decision is needed. Composite measures serve this function by assigning providers to 1 position on a scale of better-to-worse performance.

Given all these uses, NCDR believes that while we will continue to report these measures at the individual level there is a distinctive value of having a composite measure endorsed at NQF.

**1c.4. Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**

Each of the components of this measure address appropriate medication prescribing at discharge for ICD/CRT-D patients.

## 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):*

Cardiovascular

**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):*

Populations at Risk

**S.1. Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

N/A

**S.2a. If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

**S.2b. Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** icd_v2_codersdatadictionary_2-2-637061353934779116-637088191497113357.pdf

**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.

Yes

**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.

The measure language was updated to further simplify and clarify the measure intent. Specifically, the denominator was expanded to also include cardiac resynchronization therapy defibrillator (CRT-D) implant patients and the exclusions for the measure were removed since they were duplicative to what is captured and calculated in the numerator. None of these changes were substantive and do not impact the measure calculation or results.

**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).*

Generator patients who receive all medications for which they are eligible:

1. ACE/ARB prescribed at discharge (if eligible for ACE/ARB as described in denominator) AND
2. Beta blockers prescribed at discharge (if eligible for beta blockers as described in denominator)

**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).*

If eligible for ACE/ARB and given, then code “Yes”

If eligible for ACE/ARB but contraindicated, then code “No – medical reason” or "No – patient reason”

If eligible for ACE/ARB and not given, then code “No, no reason”

If eligible for beta blocker and given, then code “Yes”

If eligible for beta blocker but contraindicated, then code “No – medical reason” or "No – patient reason”

If eligible for beta blocker and not given, then code “No, no reason”

If any "No, no reason" present, then performance not met. Else, performance met.

Note: Contraindicated and those participating in blinded studies are considered performance met. There are technically no exclusions or exceptions that would remove patients from the denominator.

**S.6. Denominator Statement** *(Brief, narrative description of the target population being measured)*

All generator patients surviving hospitalization who are eligible to receive either an ACE/ARB or beta blocker at discharge.

**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).*

All generator patients surviving hospitalization who are eligible to receive any one of the two medication classes:

1) ACE/ARB: Patients who have an ejection fraction (EF) of <40%

OR

2) Beta blockers:

Patients have either:

a. EF of <40% AND/OR

b. Previous myocardial infarction (MI)

**S.8. Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

None

**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.)*

N/A

**S.10. Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

N/A

**S.11. Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

**S.12. Type of score:**

Rate/proportion

If other:

**S.13. Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

**S.14. Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

1) Check if given patient survived hospitalization and is eligible for 1 of the 2 medication therapies.

2) If eligible for at least 1 medication, then keep this patient.

3) If not eligible for any of the 2 medications, then patient is removed from eligibility.

If eligible for ACE/ARB and given, then code “Yes”

If eligible for ACE/ARB and not given, then code “No, no reason”

If eligible for ACE/ARB but contraindicated, then code “No – medical reason” or "No – patient reason”

If eligible for Beta Blocker and given, then code then “Yes”

If eligible for Beta Blocker and not given, then code “No, no reason”

If eligible for Beta Blocker but contraindicated, then code “No – medical reason” or "No – patient reason”

4) If any “No, no reason” present, then performance not met. Else, performance met.

Although ineligible cases are removed from the denominator population for the performance calculation, the number of patients with valid exceptions should be calculated and reported along with performance rates to track variations in care and highlight possible areas of focus for QI.

If the patient does not meet the numerator and a valid exception is not present, this case represents a quality failure.

Missing data defaults to “performance not met” This measure assumes that missing documentation on the process results in a failure of meeting an evidence based therapy.

**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.*

Registry Data

**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.

National Cardiovascular Data Registry (NCDR) ICD Registry

**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)*

Available in attached appendix at A.1

**S.20. Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility

**S.21. Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Inpatient/Hospital

If other:

**S.22. COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

N/A

**2. Validity – See attached Measure Testing Submission Form**

0965_composite_testing_attachment_11.21.2019.docx

**2.1 For maintenance of endorsement**

*Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

Yes

**2.2 For maintenance of endorsement**

*Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.*

No

**2.3 For maintenance of endorsement**

*Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.*

No - This measure is not risk-adjusted

## Measure Testing (subcriteria 2a2, 2b1-2b6)

**Measure Number** (*if previously endorsed*)**:** 965
**Composite Measure Title**: Discharge Medications (ACE/ARB) and beta blockers) in Eligible ICD/CRT-D Implant Patients
**Date of Submission**: Click here to enter a date

**Composite Construction:**

☐Two or more individual performance measure scores combined into one score

☒ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

### 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 different components in the composite, indicate the component after the checkbox. 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: Click here to describe | ☐ other: Click here to describe |

**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*).

2019 Submission

We analyzed data from the National Cardiovascular Data Registry's ICD Registry. This is a national quality improvement registry used in 1732 US hospitals. Rigorous quality standards are applied to the data and both quarterly and performance reports are generated for participating centers to track and improve their performance.

2015 submission

We propose to use the National Cardiovascular Data Registry for ICD Registry. This is a national quality improvement registry used in >1700 US hospitals. Some states and healthcare systems mandate participation, and participation is required as a condition for hospital reimbursement for Medicare beneficiaries receiving ICD therapy for primary prevention of sudden death. Rigorous quality standards are applied to the data and both quarterly and performance reports are generated for participating centers to track and improve their performance.

**1.3. What are the dates of the data used in testing**?

2019 submission

The NCDR ICD Registry data used for this application are reflective of a two-year period, comprising discharges between January 1, 2017 to December 31, 2018. There were no substantive changes to the Data Collection Form (version 2.2) during this period of time.

2015 submission

We have chosen to use different datasets to provide support for different aspects of the proposed measure.

Assessment of item-level reliability through the Audit Program: 01/2010-12/2010

All other forms of reliability testing: Jan 2013-Jun 2014

Hospital information about the Safety Net Hospital and % Medicaid are derived from AHA 2010 data.

**1.4. What levels of analysis were tested**? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

| **Measure Specified to Measure Performance of:** (***must be consistent with levels entered in item S.20***) | **Measure Tested at Level of:** |
|---|---|
| ☐ individual clinician | ☐ individual clinician |
| ☐ group/practice | ☐ group/practice |
| ☒ hospital/facility/agency | ☒ hospital/facility/agency |

| | |
|---|---|
| ☐ health plan | ☐ health plan |
| ☐ other: Click here to describe | ☐ other: Click here to describe |

**1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

2019 submission

The overall measured entities, following the application of exclusion criteria, are as follows:

**Table 1: Entities Evaluated by Level of Analysis**

| Level of Analysis | Variable | Data Source | Number |
|---|---|---|---|
| Patient | Patient Hospital Stay | NCDR ICD Registry | 225,665 |
| Hospital | Facilities | NCDR ICD Registry | 1,732 |

2015 submission

For all the descriptive statistics for this measure except auditing:
Number of the measured entities (hospitals): 1,606

*Assessment of item-level reliability through the Audit Program:*
To assess inter-rater reliability of the extracted data elements that comprise this measure, data from 25 participating hospitals were reviewed by an independent contractor hired by ACCF.

**1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

2019 submission

For all descriptive statistics, we used data collected by the ICD Registry between January 2017 and December 2018. Patient and hospital characteristics are presented below (Table 2).

**Table 2: Selected Characteristics by Calendar Year**

| Description | Total | | Year | | | | P |
|---|---|---|---|---|---|---|---|
| | | | 2017 | | 2018 | | |
| | # | % | # | % | # | % | |
| ALL | 22,5665 | 100.00 | 123,881 | 100.00 | 101,784 | 100.00 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Age≥65** | | | | | | | <0.001 |
| No | 83,512 | 37.01 | 45,380 | 36.63 | 38,132 | 37.46 | |
| Yes | 142,153 | 62.99 | 78,501 | 63.37 | 63,652 | 62.54 | |
| **Female** | | | | | | | <0.5067 |
| No | 166,174 | 73.64 | 91,292 | 73.69 | 74,882 | 73.57 | |
| Yes | 59,491 | 26.36 | 32,589 | 26.31 | 26,902 | 26.43 | |
| **Race** | | | | | | | <0.0001 |
| Hispanic | 16,403 | 7.27 | 8,844 | 7.14 | 7,559 | 7.43 | |
| White non-Hispanic | 167,961 | 74.43 | 92,914 | 75.00 | 75,047 | 73.73 | |
| Black non-Hispanic | 34,958 | 15.49 | 18,795 | 15.17 | 16,163 | 15.88 | |
| Other | 6,343 | 2.81 | 3,328 | 2.69 | 3,015 | 2.96 | |
| **Hospital % Non-White by Quartile** | | | | | | | <0.0001 |
| Q1 (0.00% to 4.57%) | 35,076 | 15.54 | 19,601 | 15.82 | 15,475 | 15.20 | |
| Q2 (>4.57% to 12.70%) | 63,602 | 28.18 | 35,360 | 28.54 | 28,242 | 27.75 | |
| Q3 (>12.70% to 27.20%) | 70,016 | 31.03 | 37,989 | 30.67 | 32,027 | 31.47 | |
| Q4 (>27.20%) | 56,971 | 25.25 | 30,931 | 24.97 | 26,040 | 25.58 | |
| **Hospital % Dual Medicare & Medicaid by Quartile** | | | | | | | |
| Q1 (0.00% to 0.007%) | 18,918 | 8.38 | 11,064 | 8.93 | 78,54 | 7.72 | |
| Q2 (>0.00% to 4.28%) | 78,988 | 35.00 | 42,439 | 34.26 | 36,549 | 35.91 | |
| Q3 (>4.28% to 9.36%) | 76,278 | 33.80 | 42,315 | 34.16 | 33,963 | 33.37 | |
| Q4 (>9.36%) | 51,481 | 22.81 | 28,063 | 22.65 | 23,418 | 23.01 | |
| **Met the Discharge Composite Measure** | | | | | | | <0.0001 |
| No | 28,829 | 12.78 | 16,330 | 13.18 | 12,499 | 12.28 | |
| Yes | 196,836 | 87.22 | 107,551 | 86.82 | 89,285 | 87.72 | |

---

2015 submission

The number of patients varies by testing type.

For all the descriptive statistics for this measure except auditing we used data submitted to the ICD Registry between January 2013 and June 2014. Note this reflects all data from all centers that met data quality standards irrespective of the case volume of participating hospitals.

*Selected Characteristics by Calendar Year*

| Description | Total | Year | |
|---|---|---|---|
| | | Jan – Dec 2013 | Jan – Jun 2014 |

| | # | % | # | % | # | % |
|---|---|---|---|---|---|---|
| ALL | 195563 | 100.00 | 131193 | 100.00 | 64370 | 100.00 |
| Age>=65 | | | | | | |
| No | 70743 | 36.17 | 47084 | 35.89 | 23659 | 36.75 |
| Yes | 124820 | 63.83 | 84109 | 64.11 | 40711 | 63.25 |
| Female | | | | | | |
| No | 145765 | 74.54 | 97732 | 74.49 | 48033 | 74.62 |
| Yes | 49798 | 25.46 | 33461 | 25.51 | 16337 | 25.38 |
| RACE | | | | | | |
| Hispanic | 11268 | 5.76 | 7541 | 5.75 | 3727 | 5.79 |
| White non-hispanic | 152042 | 77.75 | 102370 | 78.03 | 49672 | 77.17 |
| Black non-Hispanic | 27925 | 14.28 | 18421 | 14.04 | 9504 | 14.76 |
| Other | 4328 | 2.21 | 2861 | 2.18 | 1467 | 2.28 |
| Safety Net Hospital* | | | | | | |
| Unknown | 2342 | 1.20 | 1588 | 1.21 | 754 | 1.17 |
| No | 164694 | 84.22 | 110503 | 84.23 | 54191 | 84.19 |
| Yes | 28527 | 14.59 | 19102 | 14.56 | 9425 | 14.64 |
| Hospital % Non-White | | | | | | |
| Q1 (0.00% to 3.16%) | 27378 | 14.00 | 18254 | 13.91 | 9124 | 14.17 |
| Q2 (>3.16% to 10.57%) | 56591 | 28.94 | 37975 | 28.95 | 18616 | 28.92 |
| Q3 (>10.57% to 24.12%) | 64411 | 32.94 | 43568 | 33.21 | 20843 | 32.38 |
| Q4 (>24.12%) | 47183 | 24.13 | 31396 | 23.93 | 15787 | 24.53 |
| Hospital % Medicaid* | | | | | | |
| Unknown | 2342 | 1.20 | 1588 | 1.21 | 754 | 1.17 |
| Q1 (0.00% to 12.70%) | 50024 | 25.58 | 33968 | 25.89 | 16056 | 24.94 |
| Q2 (>12.70% to 18.41%) | 52577 | 26.88 | 34940 | 26.63 | 17637 | 27.40 |
| Q3 (>18.41% to 22.72%) | 49841 | 25.49 | 33299 | 25.38 | 16542 | 25.70 |
| Q4 (>22.72%) | 40779 | 20.85 | 27398 | 20.88 | 13381 | 20.79 |
| Met the Composite Measure | | | | | | |
| No | 36242 | 18.53 | 24699 | 18.83 | 11543 | 17.93 |
| Yes | 159321 | 81.47 | 106494 | 81.17 | 52827 | 82.07 |

* Hospital information about the Safety Net Hospital and % Medicaid are derived from AHA 2010 data.

*Assessment of item-level reliability through the Audit Program:*
To assess inter-rater reliability of the extracted data elements that comprise this measure, we reviewed

627 patients.

**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**.

2019 submission

The datasets, dates, number of measured entities, and number of admissions for all forms of reliability and validity testing were from an uninterrupted 2-year period: 01/2017 to 12/2018.

2015 submission

There are different time periods and different descriptive statistics as noted in previous sections. The datasets, dates, number of measured entities, and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2) using audit data: 01/2010 – 12/2010
For the split sample testing: 01/2013 – 06/2014

**1.8 What were the social risk factors that were available and analyzed?** For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

2019 submission

This measure examines compliance with processes of care performed at the time of discharge and it does not require risk adjustment. We examined age, gender, race/ethnicity, proportion of non-white patients, and proportion of patients who have dual Medicare/Medicaid eligibility to determine if there were differences in these demographic indicators of social risk as discussed in section 2b4 (Identification of statistically significant and meaningful differences in performance.)

2015 submission

We do not currently collect many of the SDS variables examples listed above. However, we do collect data on race as well as insurance type.

---

**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*)

***Note****: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*

☒ **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. Describe the method of reliability testing and what it tests** (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

2019 submission

Similar to our 2015 submission, we used a split sample methodology.

For the performance rates and social risk data, unadjusted rates were calculated and a Pearson correlation coefficient and ICC was computed.

2015 submission

*Split Sample Methodology*
For the performance rates and disparities data, raw rates were calculated and a correlation coefficient was computed.

*Assessment of item-level reliability through the Audit Program:*
To assess inter-rater reliability of the extracted data elements that comprise this measure, 627 patients at 25 hospitals were reviewed by an independent contractor hired by ACCF.

Assessment of item-level reliability through the Audit Program:
The NCDR Data Quality Program ensures that data submitted to the NCDR are collected completely and in a valid manner. The NCDR Data Quality Program consists of 3 main components: data completeness, consistency, and accuracy. Completeness focuses on the proportion of missing data within fields, whereas consistency determines the extent to which logically related fields contain values consistent with other fields. Accuracy characterizes the agreement between registry data and the contents of original charts from the hospitals submitting data. Before entering the Enterprise Data Warehouse (EDW), all submissions are scored for file integrity and data completeness, receiving 1 of 3 scores that are transmitted back to facilities using a color-coding scheme. A "red light" means that a submission has failed because of file integrity problems such as excessive missing data and internally inconsistent data. Such data are not processed or loaded into the EDW. A "yellow light" status means that a submission has passed the integrity checks but failed in completeness according to predetermined thresholds. Such data are processed and loaded into the EDW but are not included in any registry aggregate computations until corrected. Facilities are notified about data submission problems and provided an opportunity to resubmit data. Finally, a "green light" means that a submission has passed all integrity and quality checks. Such submissions are loaded to the EDW. After passing the DQR, data are loaded into a common EDW that houses data from all registries and included for all registry aggregate computations. In a secondary transaction process, data are loaded into registry-specific, dimensionally modeled data marts.
A summary of the Program is noted under Table 1.

**Table 1. Data Quality Program Overview**

| | |
|---|---|
| **Methodology** | • Nationwide program (i.e., all submitting participants in the United States)<br>• Review of data submitted the previous year<br>• Review of a subset of data elements that can rotate each year<br>• Remote review of data combined with couple of onsite visit<br>• Onsite visits are targeted based on the Data Outlier Program<br>• Random selection of sites and records<br>• Blinded data abstraction from medical charts<br>• Inter-rater Reliability Assessment conducted to validate the audit findings<br>• Adjudication step for participant to refute audit findings |
| **Scope** | • Review of hospital's medical records for related episodes of care<br>• Assessment of complete submission (Comparison of two lists : hospital list of cases with specific billing codes versus NCDR submitted records) |
| **Criteria for selecting sites/records** | Remote audit :<br>• Sites passing their quarterly DQR for 2 quarters within audited year<br>• Sites submitting at least the number of records/sites being reviewed<br>Onsite audit<br>• Sites identified with an outlier and not contacted with the data outlier program |
| **Scoring** | NCDR uses a grading system for identifying the amount of agreement or matching between the data captured during the medical record review and data submitted to the NCDR. |

**2a2.3. What were the statistical results from reliability testing**? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

2019 submission

Split Sample Methodology

The split samples were calculated during the same timeframe to mitigate confounding factors based on time differences. The cohort was split into two random samples to compare measure scores.

**Table 3: Split Sample Composition (i.e., Number and Proportion of Patients in Each Sample by Year)**

| Description | Total | | Year | | | |
|---|---|---|---|---|---|---|
| | | | 2017 | | 2018 | |
| | # | % | # | % | # | % |
| **Randomly Split Samples** | | | | | | |
| First | 113,070 | 50.11 | 62,052 | 50.09 | 51,018 | 50.12 |
| Second | 112,595 | 49.89 | 61,829 | 49.91 | 50,766 | 49.88 |

**Table 4: Distribution of Performances for the ICD Discharge Composite Measure Stratified by the Randomly Split Samples (2017)**

| Description | Randomly Split Samples | |
|---|---|---|
| | Second | First |
| N | 1638 | 1639 |
| Mean | 82.99% | 82.91% |
| Std Deviation | 19.22% | 19.39% |
| 100% Max | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% |
| 75% Q3 | 98.48% | 98.15% |
| 50% Median | 88.46% | 88.89% |
| 25% Q1 | 74.03% | 74.29% |
| 10% | 58.33% | 57.14% |
| 5% | 50.00% | 46.43% |
| 1% | 0.00% | 0.00% |
| 0% Min | 0.00% | 0.00% |

Correlation coefficient: 0.59430

ICC: 0.82457

**Figure 1. Distribution of Performance for the ICD Discharge Composite Measure Stratified by Randomly Split Samples (2017)**

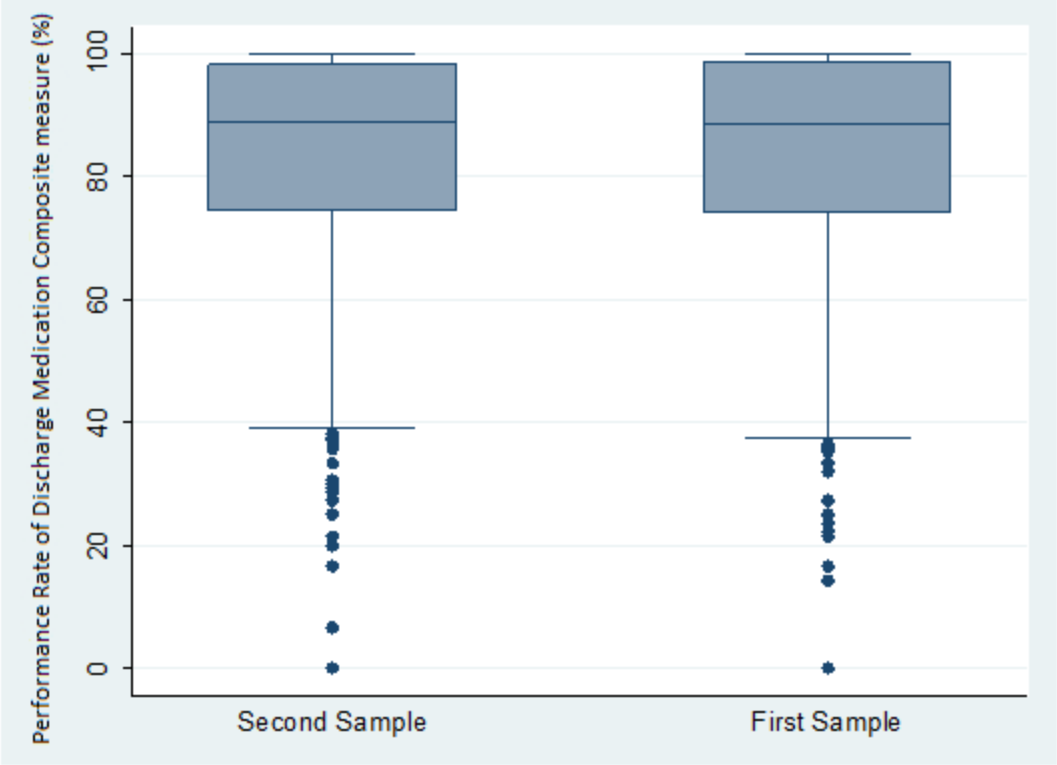


**Figure 2. Distribution of Performance for the ICD Discharge Composite Measure Stratified by Split Sample Correlation (2017)**

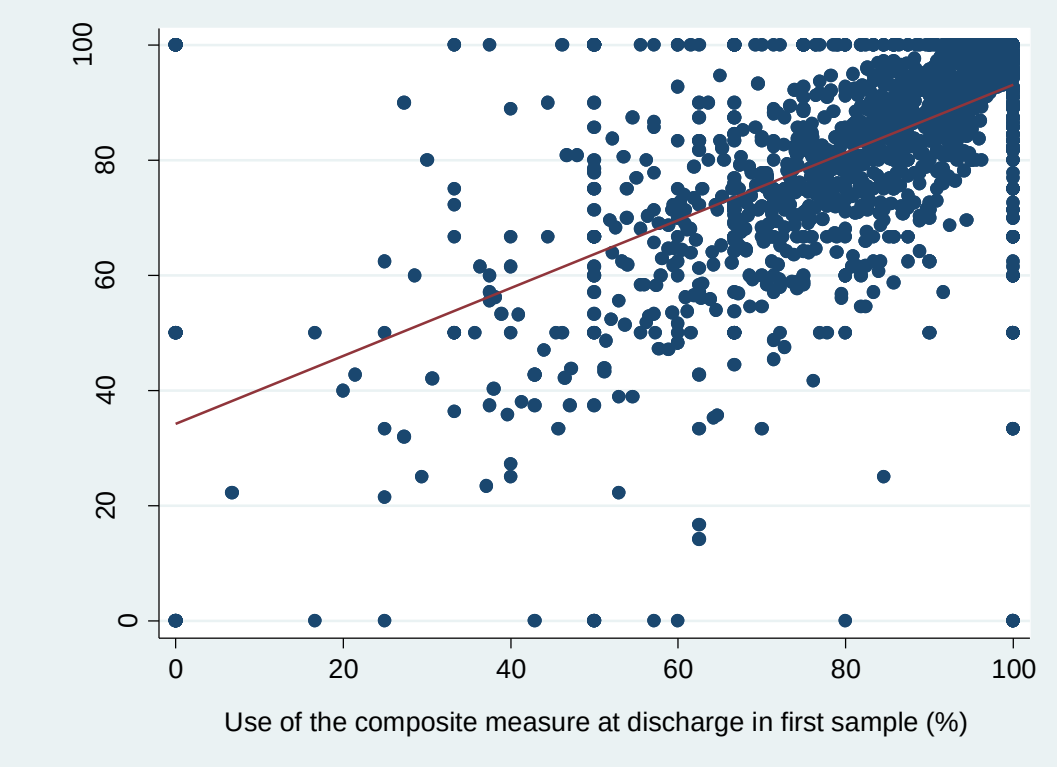



**Table 5: Distribution of Performances for the ICD Discharge Composite Measure Stratified by the Randomly Split Samples (2018)**

| | Randomly Split Samples | |
|---|---|---|
| **Description** | **Second** | **First** |
| N | 1534 | 1544 |
| Mean | 82.79% | 83.03% |
| Std Deviation | 21.26% | 21.18% |
| | | |
| 100% Max | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% |
| 75% Q3 | 100.00% | 100.00% |

| | Randomly Split Samples | |
|---|---|---|
| Description | Second | First |
| 50% Median | 90.00% | 90.00% |
| 25% Q1 | 75.00% | 75.00% |
| 10% | 55.56% | 55.56% |
| 5% | 40.00% | 40.74% |
| 1% | 0.00% | 0.00% |
| 0% Min | 0.00% | 0.00% |

Correlation coefficient: 0.52386

ICC: 0.79443

**Figure 3. Distribution of Performance for the ICD Discharge Composite Measure Stratified by Randomly Split Samples (2018)**

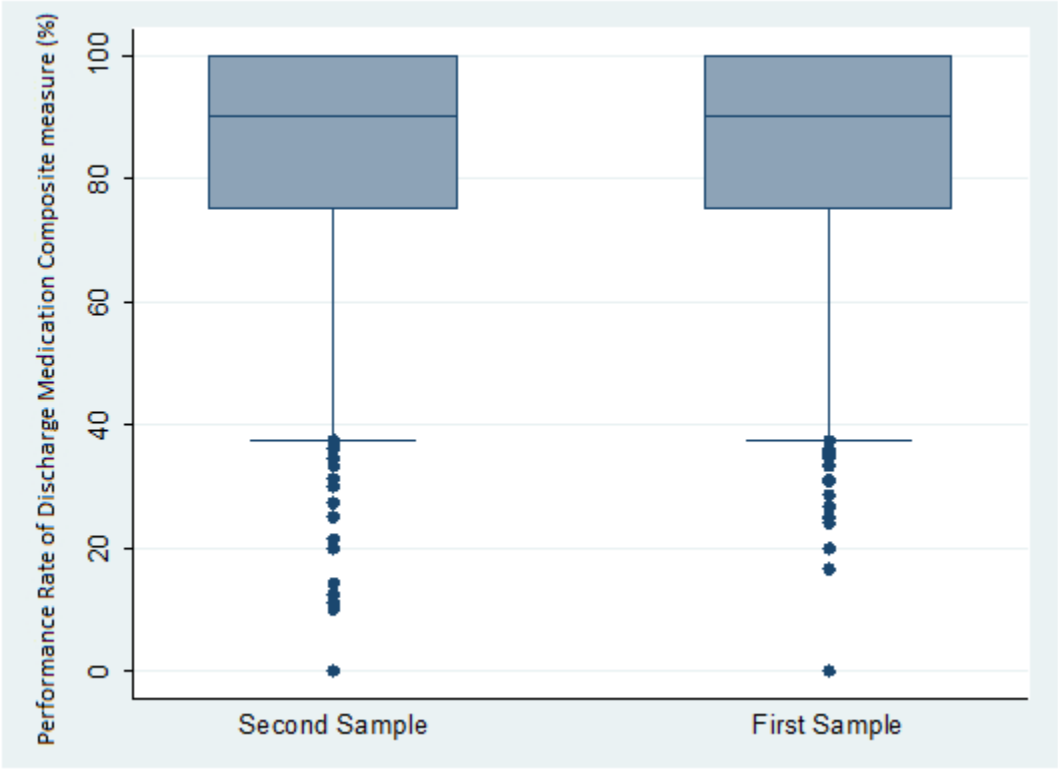
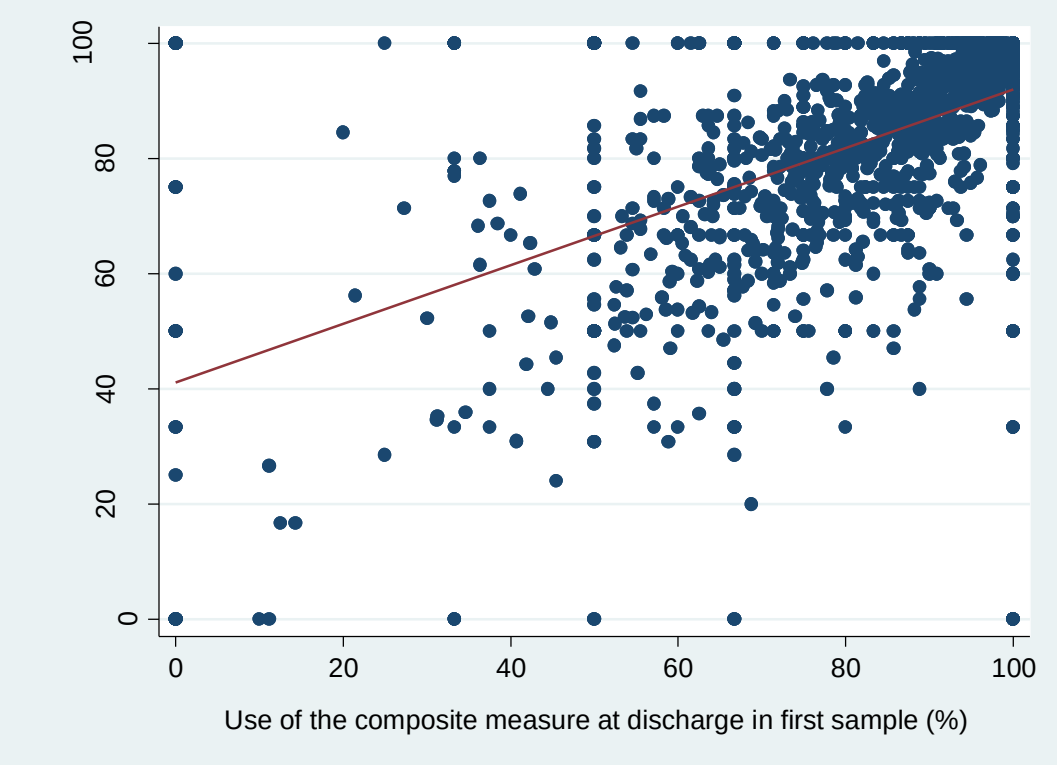


**Figure 4. Distribution of Performance for the ICD Discharge Composite Measure Stratified by Split Sample Correlation (2018)**



2015 submission

*Assessment of item-level reliability through the Audit Program:*

| CE # | field_Name | agreement rate | _KAPPA_ | L_KAPPA | U_KAPPA | N levels |
|---|---|---|---|---|---|---|
| 4170 | Prior MI | 0.815920398 | 0.60411 | 0.54118 | 0.66704 | 3 |
| 5000 | LVEF Assessed | 0.797678275 | 0.327125 | 0.245208 | 0.409041 | 3 |
| 6005 | Procedure Type | 0.978441128 | 0.955269 | 0.931311 | 0.979227 | 4 |
| 9045 | ACE Inhibitor (Any) | 0.883913765 | 0.755579 | 0.707206 | 0.803952 | 4 |
| 9100 | ARB (Any) | 0.922056385 | 0.729868 | 0.660493 | 0.799242 | 4 |
| 9110 | Beta Blocker (Any) | 0.933665008 | 0.658258 | 0.568486 | 0.748029 | 5 |

*Assessment of item-level reliability through the Audit Program:*

NCDR's Data Quality Program rotates the review of all the variables in the registry. ICD has over 300 elements that are reviewed on a 3 year rotating cycle. The elements required for this measure will be reviewed during the upcoming audit process. NCDR staff can provide kappa scores and percentage agreement scores upon completion of the cycle.

*Split Sample Methodology:*

**Distribution of hospital performance on the composite measure within random split samples (minimum 50 cases in each sample)**

| | Randomly Split Samples | |
|---|---|---|
| **Description** | **First (RAND=1)** | **Second (RAND=0)** |
| | **DCM** | **DCM** |
| N | 707 | 684 |
| Mean | 0.8178 | 0.8200 |
| Std Deviation | 0.1089 | 0.1090 |
| 100% Max | 1.0000 | 1.0000 |
| 75% Q3 | 0.9020 | 0.9087 |

| | | |
|---|---|---|
| 50% Median | 0.8199 | 0.8203 |
| 25% Q1 | 0.7414 | 0.7434 |

To evaluate the reliability of the measure, we randomly split the study cohort over the two year period (Jan 2013 to Jun 2014 combined) into two samples and restricted the cohort to hospitals that had a minimum of 50 cases in each split sample.

Results of the split sample testing are provided below. The 2 split samples were calculated during the same timeframe to avoid the potential for changes in hospital performance over time. After splitting the cohort into two random samples, we compared measure scores calculated at hospitals with at least 50 cases in both random samples. Of note, slightly less than half of participating hospitals met this volume threshold, and a few hospitals had more than 50 cases in one random sample but fewer than 50 in the other. The distribution of hospital performance was similar in the two samples (figure below), and there was an extremely high correlation between hospital performances assessed in the two samples (r 0.87949)

**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?*)

2019 submission

Split Sample Methodology

The box and whisper plot of the distribution of hospital performance for the ICD Discharge Composite Measure in 2017 and 2018 show a similar distribution of use of the composite measure at discharge for both split samples. Figures 2 and 4 show the scatterplot of the distribution of hospital performance for ICD composite measure at discharge when assessed in randomly split samples. Overall hospital performance in one random sample was strongly correlated with hospital performance in the other split sample (r=0.59430, 0.52386), for 2017 and 2018 respectively, which is consistent with a highly reliable measure.

2015 submission

*Assessment of item-level reliability through the Audit Program:*

These kappa scores were calculated with a 95% CI. By convention, a kappa > .70 is considered acceptable inter-rater reliability (Landis 1977). We used the scale below for our analysis.

0: No better than chance
0.01-0.20: Slight
0.21-0.40: Fair
0.41-0.60: Moderate
0.61-0.80: Substantial
0.81-1.0: Almost perfect

(Reference: Landis J, Koch G, The measurement of observer agreement for categorical data, *Biometrics*, 1977; 33:159-174.)

The kappa score for all medication elements demonstrate substantial or almost perfect reliability. Some of the measure elements have justifiable reasons for a lower kappa and percentage agreement scores. The element "LVEF Assessed" is not always known. Moreover, there are multiple data elements at different times during hospitalization. Therefore, it is difficult to assess which score is the correct score. Nevertheless, this element is actively discussed on monthly registry site manager calls and NCDR' s educational annual conference.

*Split Sample Methodology*

The figure above shows the scatterplot of the distribution of hospital performance for ICD composite measure at discharge when assessed in randomly split samples. Overall hospital performance in one random sample was strongly correlated with hospital performance in the other split sample (r=0.87949), which is consistent with a highly reliable measure.

---

**2b1. VALIDITY TESTING**

***Note**: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

**2b1.1. What level of validity testing was conducted**?

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Composite 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.

☐ **Validity testing for component measures** *(check all that apply)*

***Note**: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☐ **Endorsed (or submitted) as individual performance measures**

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Empirical validity testing of the component measure score(s)**

☐ **Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

**2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests** (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

2019 submission

No additional validity testing was completed.

2015 submission

*Systematic assessment of content validity:*

Content validity of this process was achieved by the specialized expertise of those individuals who developed this measure as well as the structured discussions that the group conducted. For this particular topic those individuals who were involved in identifying the key attributes and variables for this process measure were leaders and experts in the field of electrophysiology. Serial phone calls were held to both define the eligible population and given process. These clinical leaders are noted below.

NCDR Clinical Measures workgroup ensured the measure demonstrated an opportunity for improvement, had strong clinical evidence, and was a reliable and valid measure. These members included Drs. Jeptha Curtis (Chair), Frederick Masoudi, John Rumsfeld, Mark Kremers, and Matthew Reynolds.

NCDR Scientific Quality and Oversight Committee—a committee that served as the primary resource for crosscutting scientific and quality of care methodological issues. These members included Drs. Frederick Masoudi (Chair) , David Malenka, Thomas Tsai, Matthew Reynolds, David Shahian, John Windle, Fred Resnic, John Moore, Deepak Bhatt, James Tcheng, Jeptha Curtis, Paul Chan, Matthew Roe, and John Rumsfeld.

Lastly the 16 member NCDR Management Board and 31member ACCF Board of Trustees reviewed and approved these measures for submission to NQF.

*Evidence:*

ACE/ARB

ACE inhibitors reduce morbidity, mortality, and hospitalizations for patients with heart failure and left ventricular systolic dysfunction. The efficacy of ARB therapy has been strengthened by several large-scale prospective randomized clinical trials demonstrating lower rates of death and heart failure hospitalization among patients with heart failure and LVSD. Consensus clinical guidelines include strong recommendations for ACE inhibitors for all patients with HF due to LV systolic dysfunction unless they have a contraindication to their use or have been shown to be unable to tolerate treatment with these drugs. ACE inhibitors remain the first choice for inhibition of the renin-angiotensin system in chronic HF, but ARBs are considered a reasonable alternative. Even if the patient has responded favorably to the diuretic, treatment with ACE inhibitor or ARBs should be initiated and maintained in patients who can tolerate them, because they have been shown to favorably influence the long-term prognosis of HF

Beta Blocker-MI

The benefits of beta blocker therapy in patients with prior myocardial infarction without contraindications have been established for a wide range of patient groups. The greatest benefits are seen in patients with the greatest baseline risk: those with impaired ventricular function or ventricular arrhythmias and those who do not undergo reperfusion. The benefits of beta-blocker therapy for secondary prevention are well established.

Beta Blocker-LVSD

Long term beta blocker therapy for patients with left systolic ventricular dysfunction (LVSD) can improve symptoms of heart failure, improve patient clinical status, and reduce hospitalizations and mortality.
All this research demonstrates that this measure contributes to improved intermediate outcomes and important outcomes such as reductions in hospitalizations and mortality rates.

*Empiric assessment of content validity:*

As noted in the measure application, we conducted empiric analyses to assess the association of patient and hospital performance on the composite measure with adverse outcomes, specifically mortality and readmission at 6 months following hospital discharge. To conduct these analyses we used a sample of patients for whom these outcomes were available. This consisted of 93971 Medicare fee-for-service patients at least 65 years of age who underwent ICD implantation in 2010 or 2011. Our outcomes of interest included all-cause mortality, all-cause readmission, and the combination of the 2 at 6 months following hospital discharge. We examined the proportion of patients who experienced these outcomes stratified by whether or not they were discharged on appropriate medical therapy. In addition, we conducted analyses at the hospital level examining the association between hospital-level performance on the measure and the combination of mortality or readmission at 6 months.

**2b1.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test*)

2019 submission

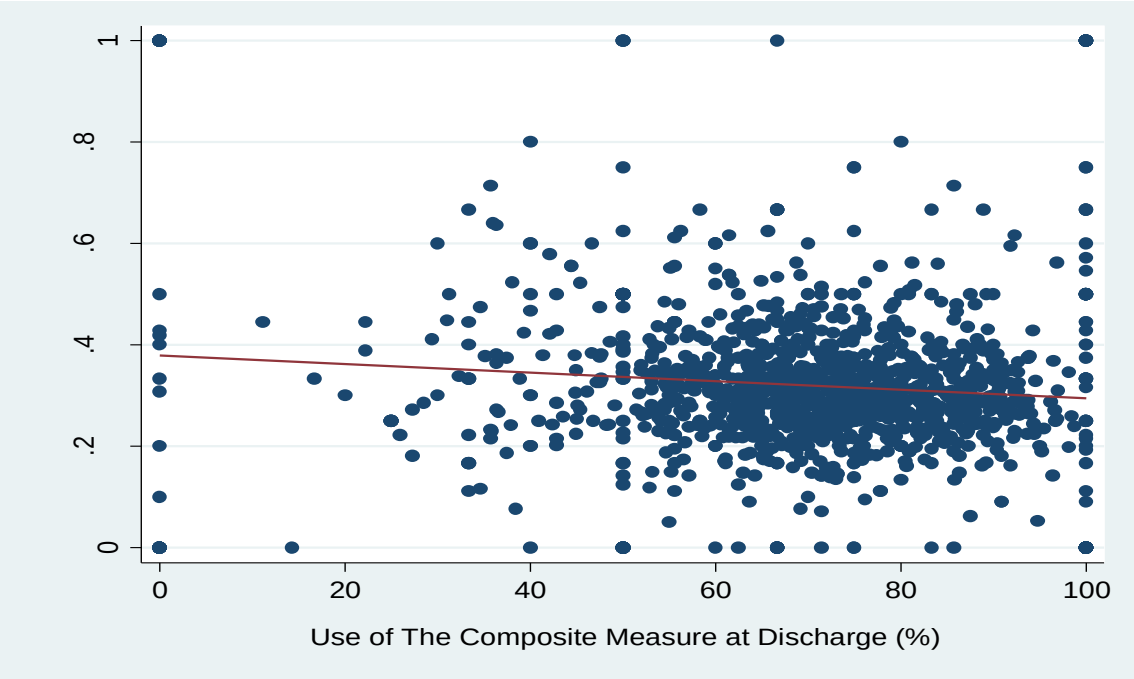
No additional validity testing was completed.

2015 submission

Patient-level results are shown below. Overall, a significantly smaller proportion of patients discharged on appropriate medical therapy died or were readmitted within 6 months of hospital discharge.

| Description | Use of Medications | | | | P |
|---|---|---|---|---|---|
| | No | | Yes | | |
| | # | % | # | % | |
| **Composite Measure** | **25217** | | **68754** | | |
| 6 month mortality | 2408 | 9.55 | 3720 | 5.41 | <0.001 |
| 6 month readmission | 8587 | 34.05 | 18643 | 27.12 | <0.001 |
| 6 month mortality or readmission | 9148 | 36.28 | 19504 | 28.37 | <0.001 |

Hospital-level results are shown below. The figure shows the association between rate of death or readmission within 6 months of discharge, with the use of the composite measure at discharge. Hospital performance on the composite discharge medication measure were significantly correlated with the combined outcome of death or readmission such that patients treated at hospitals that performed better on the measure had better unadjusted outcomes that those treated at hospitals that performed worse on the measure (correlation coefficient (-0.0998), $p<0.001$).




**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?*)

2019 submission

No additional validity testing was completed.

2015 submission

These findings support the validity of the composite discharge medication measure. At both the patient and hospital level, performance on the measure was associated with better outcomes at 6 months following discharge.

*Threats to Validity:*

**Information Bias:** There should be little concern for information bias since the care process is objective and there is a low likelihood of misreporting the given care process. Additionally, since there is only 1 data source that is used for NCDR inpatient registries thus mitigating this potential threat.

**Missing Data Bias:** Because of the large amount of data typically contained in registries, it is not feasible to meet the stringent requirements used in clinical trials. However, unlike with administrative claims data, data fields in a registry must be assessed for completeness, consistency, and accuracy to support the central activities of the registry. The NCDR Data Quality Program consists of 3 main components: data completeness, consistency, and accuracy. Completeness

focuses on the proportion of missing data within fields, whereas consistency determines the extent to which logically related fields contain values consistent with other fields. Accuracy characterizes the agreement between registry data and the contents of original charts from the hospitals submitting data. The thresholds for all critical elements in a performance measure are set high to ensure data completeness and consistency for the overall calculation of the performance measure. Therefore it is unlikely missing data bias would threaten the validity properties.

**Selection Bias:** In January 2005 the Centers for Mediare and Medicaid Services (CMS) expanded the covered indications for primary prevention ICDs to incorporate the findings from published literature. As part of this expansion, CMS mandated that a national registry be formed to compile data on Medicare patients implanted with primary prevention ICDs to confirm the appropriateness of ICD utilization in this patient population. CMS selected the NCDR ICD Registry as the mandated national registry in October 2005 and enrollment opened on January 1, 2006. As the CMS-mandated registry for hospitals that perform ICD implantation procedures, the ICD Registry essentially requires all hospitals that receive Medicare funding, to a participant of the NCDR Registry. This limits the potential for selection bias. Additionally, based on the entity and patient descriptive statistics, there does not appear to be certain subgroups of hospitals or patients who are excluded. Lastly, the exclusion frequencies did not appear to be unusually high.

**Confounding Bias:** No empirical testing was performed since this metric is neither an outcome or resource use measure.

---

**2b2. EXCLUSIONS ANALYSIS**

***Note:*** *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*

**NA ☐ no exclusions — *skip to section 2b4***

**2b2.1. Describe the method of testing exclusions and what it tests** (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2019 submission

This measure was revised to further clarify and simplify the measure calculation. The exclusions previously included in the measure language were removed since they were duplicative to what is captured and calculated in the numerator. None of these changes are substantive and do not impact the measure calculation or results. The numbers and percentages of patient stays and facilities that are removed from the measure are provided in Table 6.

2015 submission

The only exclusions for this measure are noted under S.10. (Discharge status of expired; not eligible for either ACE/ARB or beta blockers). These exclusions are relatively rare and firmly supported by the clinical rationale.

**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*)

2019 submission

**Table 6. Inclusions/Exclusions**

| Inclusion/Exclusion | Patient Stays | | Facilities | |
|---|---|---|---|---|
| | N | % | N | % |
| **Total** | **427,995** | **100.0** | **1,787** | **100.0** |
| Discharge not in 2017 and 2018 | 128,176 | 29.9 | 44 | 2.5 |
| **Remaining** | **299,819** | **70.1** | **1,743** | **97.5** |
| Died during hospital | 1,031 | 0.3 | 0 | 0.0 |
| **Remaining** | **298,788** | **99.7** | **1,743** | **100.0** |
| Not eligible to the composite measure | 73,123 | 24.5 | 11 | 0.6 |
| **Measure Cohort** | **225,665** | **75.5** | **1,732** | **99.4** |
| The composite measure at discharge | 196,836 | 87.22 | 1,714 | 98.96 |

2015 submission

| Exclusions | Patient Stays | | Facilities | |
|---|---|---|---|---|
| **Total** | **665983** | **100.0** | **1709** | **100.0** |
| Discharge not in 2013 and 2014 | 420784 | 63.2 | 96 | 5.6 |
| **Remaining** | **245199** | **36.8** | **1613** | **94.4** |
| Died during hospital | 710 | 0.3 | 0 | 0.0 |
| **Remaining** | **244489** | **99.7** | **1613** | **100.0** |
| Not eligible to the composite measure | 48926 | 20.0 | 7 | 0.4 |
| **Study Cohort** | **195563** | **80.0** | **1606** | **99.6** |
| The composite measure at discharge | 159321 | 81.47 | 1589 | 98.94 |

**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*)

2019 submission

While the measure language has been revised to clarify and simplify the measure calculation, the results in Table 6 when compared to the table from 2015 show that similar numbers and percentages of patient stays and facilities remain in the denominator or measure cohort. Specifically, 75.5% of patient stays and 99.4% of all facilities are included based on the most recent analysis, which is only a difference of -4.5% and -0.2% from the 2015 analysis, respectively.

2015 submission

As noted above, there are no 'discretionary' exclusions. All exclusions are necessary to the accurate calculation of performance on the composite measure. For example, patients need to survive to discharge to be eligible for the measure. Similarly, it would be inappropriate to calculate the measure among patients ineligible for the medications.

_____

**2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES**
***Note:*** *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*
***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?** *(check all that apply)*

☐ **Endorsed (or submitted) as individual performance measures**

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

**2b3.1.1 If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.**

**2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities**.

**2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk** (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion;** for example, are social risk factors added after all clinical factors?

**2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:**

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

**2b3.4a. What were the statistical results of the analyses used to select risk factors?**

**2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors** (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.) Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.

**2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach** (*describe the steps—do not just name a method; what statistical analysis was used*)

*Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below*.
***If stratified, skip to 2b3.9***

**2b3.6. Statistical Risk Model Discrimination Statistics** (*e.g., c-statistic, R-squared*)**:**

**2b3.7. Statistical Risk Model Calibration Statistics** (*e.g., Hosmer-Lemeshow statistic*):

**2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves**:

**2b3.9. Results of Risk Stratification Analysis**:

**2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)?** (i.*e., what do the results mean and what are the norms for the test conducted*)

**2b3.11. Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

---

**2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE**

***Note:*** *Applies to the composite performance measure.*

**2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified** (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2019 submission

We examined variation in hospital performance for the composite measure based on overall performance, and stratified by subgroups of sex, age, and race/ethnicity and dual eligibility for Medicare and Medicaid to identify if there were meaningful differences in social risk.

2015 submission

We examined variation in hospital performance for the composite measure based on sex, age, race, and the proportion of patients who are insured through Medicaid to identify meaningful differences.

**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*)

2019 submission

**Overall**

The median rate of performance of the discharge medications composite across all hospitals in 2017 was 87.6%. There was variation in performance ranging from 74.3% to 96.8% for the first and third quartile of hospitals, respectively (Table 7 ), and the distribution was left-skewed such that the majority of hospitals scored between 80% to 100% on the ICD Discharge Measure (Figure 5 ).

In 2018, the median rate of performance of the discharge medications composite across all hospitals was 88% (IQR: 75% to 97.4%). The distribution was also left-skewed such that the majority of hospitals scored between 80% to 100% on the ICD Discharge Measure (Figure 6).

**Table 7. Distribution of Performance of the ICD Discharge Composite Measure and its Components (2017)**

| Description | Discharge Composite Measure (%) | ACEI/ARB (%) | Beta Blockers (%) |
|---|---|---|---|
| N | 1674 | 1666 | 1674 |
| Mean | 82.72% | 83.69% | 93.74% |
| Std Deviation | 18.00% | 17.54% | 10.32% |
| 100% Max | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% | 100.00% |
| 75% Q3 | 96.84% | 97.22% | 100.00% |

| | | | |
|---|---|---|---|
| 50% Median | 87.56% | 88.24% | 96.88% |
| 25% Q1 | 74.29% | 75.00% | 91.67% |
| 10% | 60.00% | 61.90% | 85.19% |
| 5% | 48.39% | 50.00% | 77.78% |
| 1% | 0.00% | 0.00% | 50.00% |
| 0% Min | 0.00% | 0.00% | 0.00% |

**Figure 5. Histogram of Hospital Performance of the ICD Discharge Composite Measure (2017)**

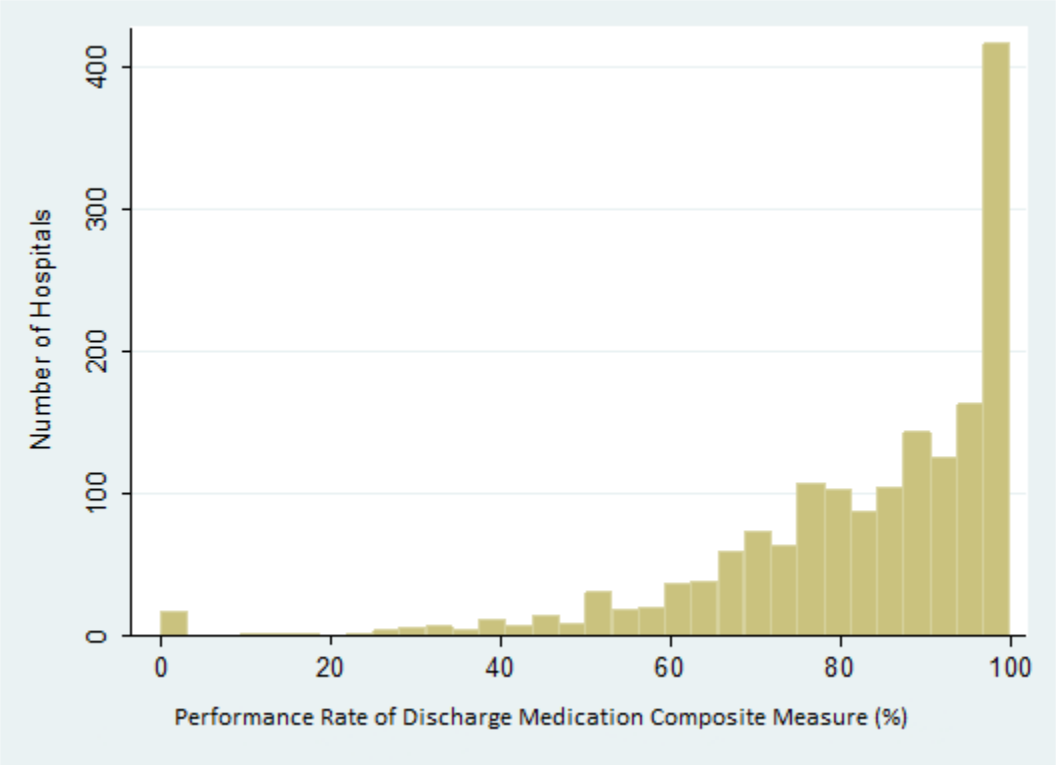
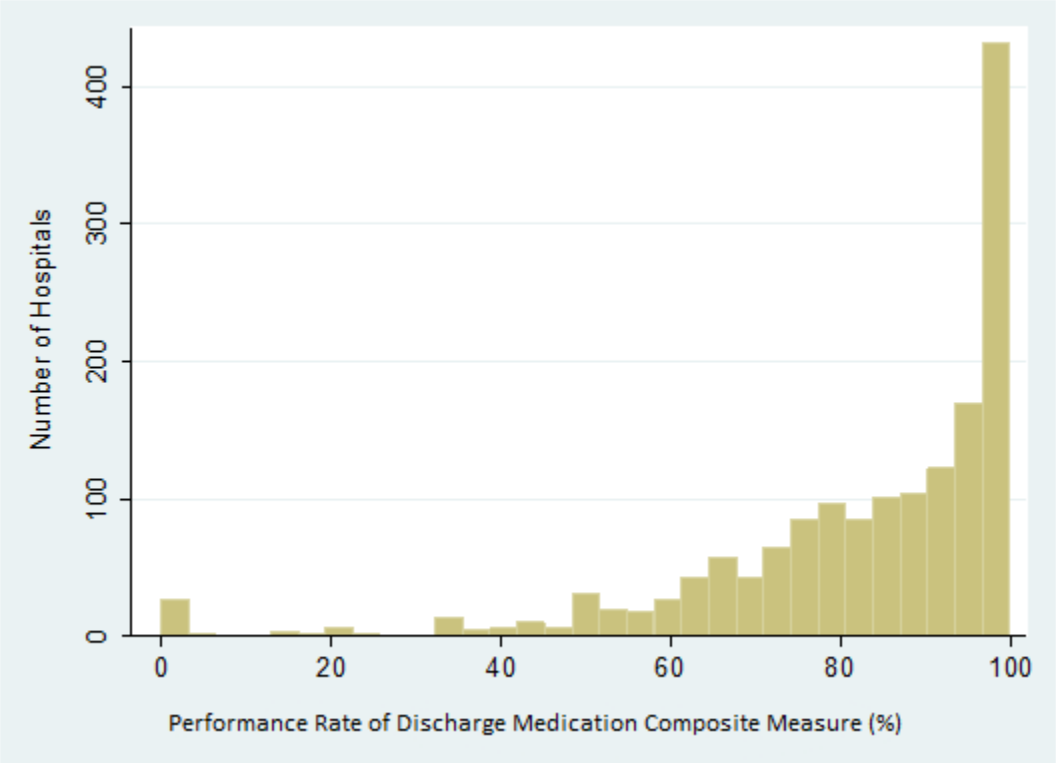



**Table 8. Distribution of Performance of the ICD Discharge Composite Measure and its Components (2018)**

| Description | Discharge Composite Measure (%) | ACEI/ARB (%) | Beta Blockers (%) |
|---|---|---|---|
| N | 1574 | 1564 | 1573 |

| | | | |
|---|---|---|---|
| Mean | 82.54% | 83.94% | 93.34% |
| Std Deviation | 19.80% | 18.68% | 12.45% |
| | | | |
| 100% Max | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% | 100.00% |
| 75% Q3 | 97.41% | 97.92% | 100.00% |
| 50% Median | 87.95% | 89.32% | 97.20% |
| 25% Q1 | 75.00% | 76.15% | 91.67% |
| 10% | 59.80% | 62.50% | 84.00% |
| 5% | 46.67% | 50.00% | 76.19% |
| 1% | 0.00% | 0.00% | 25.00% |
| 0% Min | 0.00% | 0.00% | 0.00% |

**Figure 6. Histogram of Hospital Performance of the ICD Discharge Composite Measure (2018)**



**Subgroups**

Across stratified analyses based on sex, age, race, proportion of patients who are non-White, and proportion of patients who have dual eligibility, we found variation in the distribution of hospital performance, as detailed below.

***Proportion of Non-White***

Hospitals (N=1,732) were stratified into quartiles by the proportion of non-White patients. In 2017, the median performance for those hospitals with the fewest non-white patients (Q1) was 88.2% (IQR: 71.4% to 100%). Among those hospitals with the highest proportion of non-White patients (Q4), the median performance was 87.5% (IQR: 73.5% to 96.6%).

In 2018 (Table 10), the median performance for those hospitals with the fewest non-white patients (Q1) was 85.2% (IQR: 66.7% to 99%). Among those hospitals with the highest proportion of non-White patients (Q4), the median performance was 88.5% (IQR: 75% to 97.2%) Overall, hospitals with varying proportions of non-White patients perform similarly for the ICD Discharge Composite Measure.

**Table 9. Distribution of the ICD Discharge Composite Measure Stratified by Hospital Quartile Non-White at the Hospital Level 2017 (N=1,732)**

| Description | Non-White (%) | | | |
|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 |
| Mean | 81.29% | 85.46% | 82.03% | 82.10% |
| Std Deviation | 22.11% | 13.97% | 15.62% | 18.99% |
| 100% Max | 100.00% | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 99.39% | 99.48% | 100.00% |
| 75% Q3 | 100.00% | 96.67% | 95.24% | 96.55% |
| 50% Median | 88.23% | 88.89% | 84.29% | 87.50% |
| 25% Q1 | 71.43% | 78.95% | 73.61% | 73.53% |
| 10% | 52.17% | 68.24% | 61.54% | 57.14% |
| 5% | 37.50% | 59.26% | 50.00% | 46.53% |
| 1% | 0.00% | 37.84% | 30.56% | 0.00% |
| 0% Min | 0.00% | 11.11% | 25.00% | 0.00% |

| Description | Non-White (%) | | | |
|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 |
| Median scores Test p=0.0095 | | | | |

**Figure 7. Distribution of Performance for the ICD Discharge Composite Measure Stratified by Quartiles of Non-White Patients at the Hospital Level (2017)**

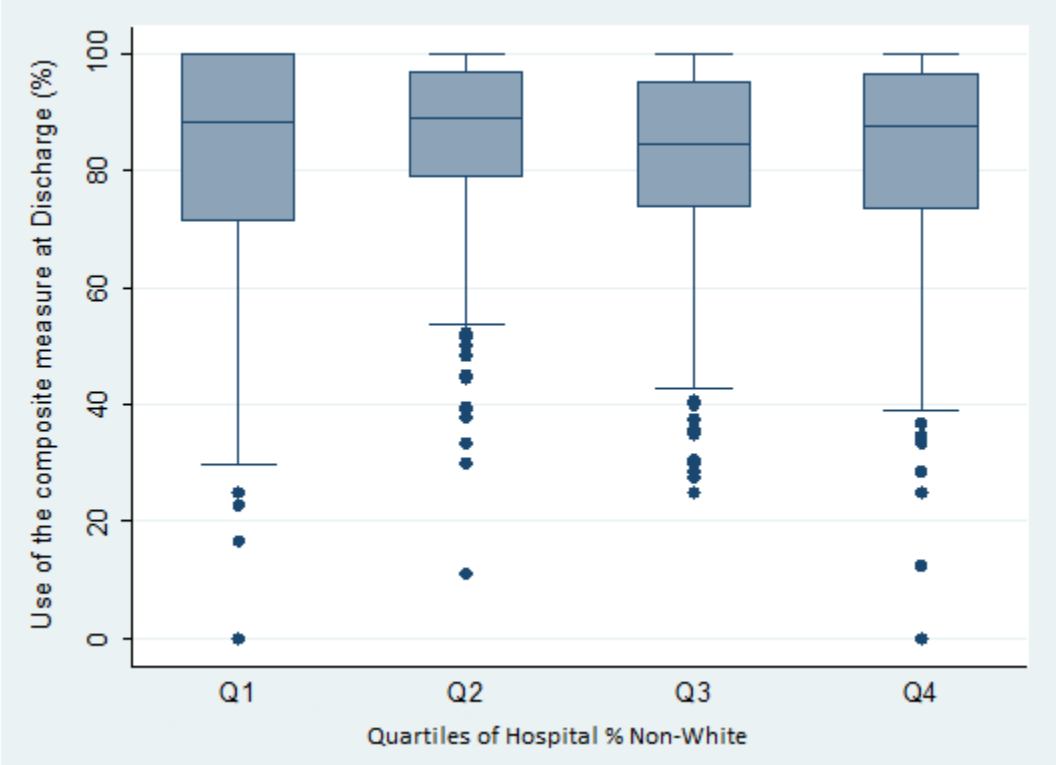
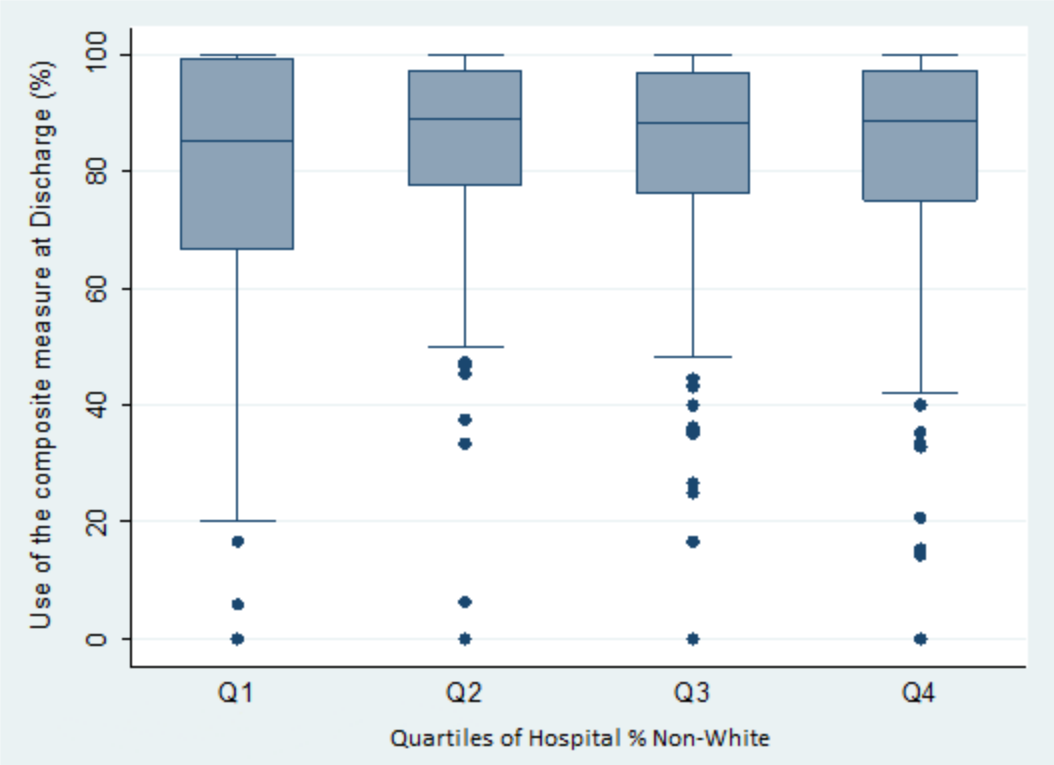



**Table 10. Distribution of the ICD Discharge Composite Measure Stratified by Hospital Quartile Non-White at the Hospital Level 2018 (N=1,732)**

| Description | Non-White (%) | | | |
|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 |
| Mean | 77.67% | 85.87% | 84.16% | 82.48% |
| Std Deviation | 25.81% | 14.34% | 16.08% | 20.07% |
| | | | | |
| 100% Max | 100.00% | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% | 100.00% | 100.00% |
| 75% Q3 | 98.98% | 97.22% | 96.67% | 97.22% |
| 50% Median | 85.24% | 89.02% | 88.14% | 88.46% |
| 25% Q1 | 66.67% | 77.78% | 76.27% | 75.00% |
| 10% | 45.00% | 66.67% | 62.16% | 55.56% |
| 5% | 5.88% | 61.11% | 53.85% | 42.11% |
| 1% | 0.00% | 37.50% | 25.00% | 0.00% |
| 0% Min | 0.00% | 0.00% | 0.00% | 0.00% |

Median scores Test p=0.0908

**Figure 8. Distribution of Performance for the ICD Discharge Composite Measure Stratified by Quartiles of Non-White Patients at the Hospital Level (2018)**




***Gender***

In 2017, the median hospital performance among female patients was 90% (IQR: 75% to 100%). Among male patients, the performance median was 88% (IQR: 75% to 97.3%) (Table 11, Figure 9).

In 2018, the median hospital performance among female patients was 91.4% (IQR: 75% to 100%). Among male patients, the performance median was 88.9% (IQR: 75% to 98%) (Table 12, Figure 10).

The P-values of 0.0148 and 0.0119 for 2017 and 2018 respectively, indicats that there is a statistically significant difference in median performance of the ICD Discharge Composite measure between male and female ICD patients.Men experience more variation in provision of ICD discharge medications and have a slightly lower median performance rate compared to women.

**Table 11. Distribution of Performance Rate for the ICD Discharge Composite Measure Stratified by Gender at the Hospital-Level 2017 (N=1,732)**

| | Gender | |
|---|---|---|
| **Description** | **Male** | **Female** |
| Mean | 82.88% | 82.86% |
| Std Deviation | 18.50% | 21.45% |

| | | |
|---|---|---|
| 100% Max | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% |
| 75% Q3 | 97.30% | 100.00% |
| 50% Median | 88.00% | 90.00% |
| 25% Q1 | 75.00% | 75.00% |
| 10% | 60.24% | 50.00% |
| 5% | 50.00% | 40.00% |
| 1% | 0.00% | 0.00% |
| 0% Min | 0.00% | 0.00% |

Median Scores Test P=0.0148

**Figure 9. Distribution of Performance of the ICD Discharge Composite Measure Stratified by Gender at the Hospital-Level 2017**

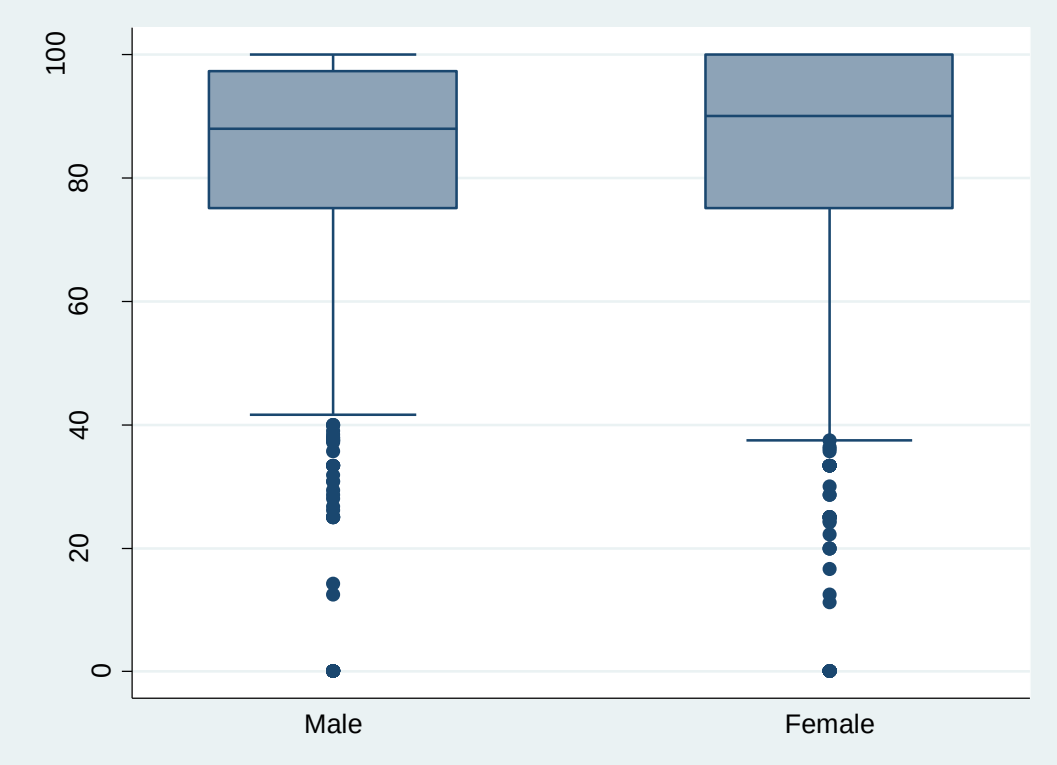
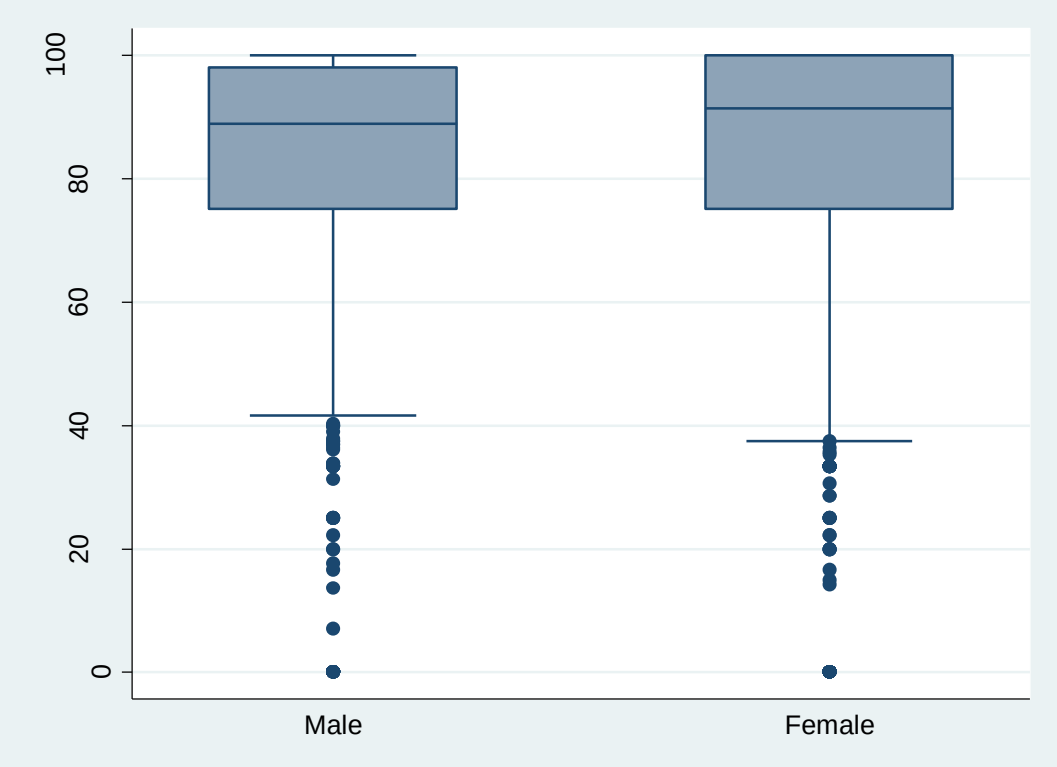



**Table 12. Distribution of Performance Rate for the ICD Discharge Composite Measure Stratified by Gender at the Hospital-Level 2018 (N=1,732)**

| Description | Gender |
|---|---|

| | Male | Female |
|---|---|---|
| Mean | 82.84% | 83.43% |
| Std Deviation | 20.00% | 22.90% |
| | | |
| 100% Max | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% |
| 75% Q3 | 98.04% | 100.00% |
| 50% Median | 88.89% | 91.43% |
| 25% Q1 | 75.00% | 75.00% |
| 10% | 58.14% | 52.94% |
| 5% | 47.62% | 33.33% |
| 1% | 0.00% | 0.00% |
| 0% Min | 0.00% | 0.00% |
| | | |
| Median Scores Test p=0.0119 | | |

**Figure 10. Distribution of Performance of the ICD Discharge Composite Measure Stratified by Gender at the Hospital-Level 2018**




***Age***

In 2017, the median hospital performance among patients aged < 65 was 91.04% (IQR: 76.9% to 100%) and the median hospital performance for patients ≥ 65 years of age was 86.7% (IQR: 72.7% to 97.5%) (Table 13, Figure 11).

In 2018, the median hospital performance among patients aged <65 was 93% (IQR: 78% to 100%). The median hospital performance for patients ≥ 65 years of age was 88% (IQR: 73% to 98.3%) Table 14, Figure 12. The P-values indicate that that the median performance between the two groups is statistically significant, with older patients experiencing a lower rate of provision of ICD discharge medications than younger patients.

**Table 13. Distribution of the Performance of the ICD Discharge Composite Measure Stratified by Age at the Hospital-Level 2017 (N=1,732)**

| Description | Age < 65 | Age ≥ 65 |
|---|---|---|
| Mean | 84.18% | 81.99% |

| | | |
|---|---|---|
| Std Deviation | 20.33% | 19.25% |
| | | |
| 100% Max | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% |
| 75% Q3 | 100.00% | 97.50% |
| 50% Median | 91.04% | 86.96% |
| 25% Q1 | 76.92% | 72.73% |
| 10% | 58.06% | 58.82% |
| 5% | 50.00% | 44.44% |
| 1% | 0.00% | 0.00% |
| 0% Min | 0.00% | 0.00% |

Median Scores test p < .0001

**Figure 11. Distribution of Performance of the Composite Measure at Discharge Stratified by Age Group at the Hospital-Level 2017**

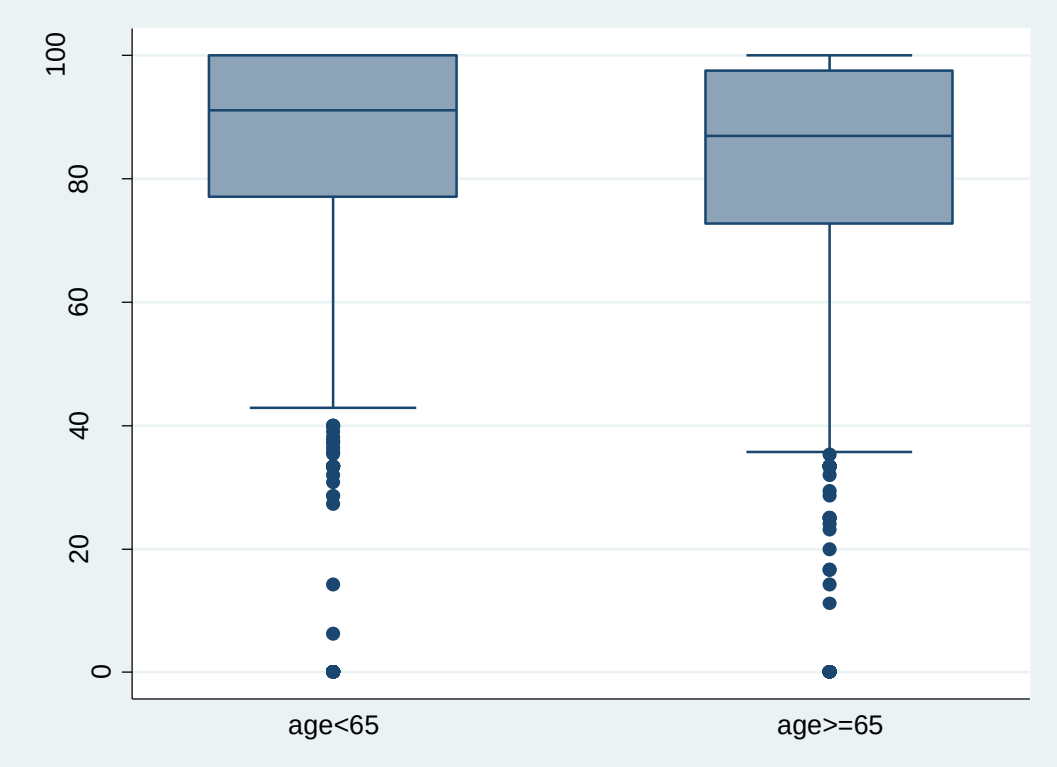
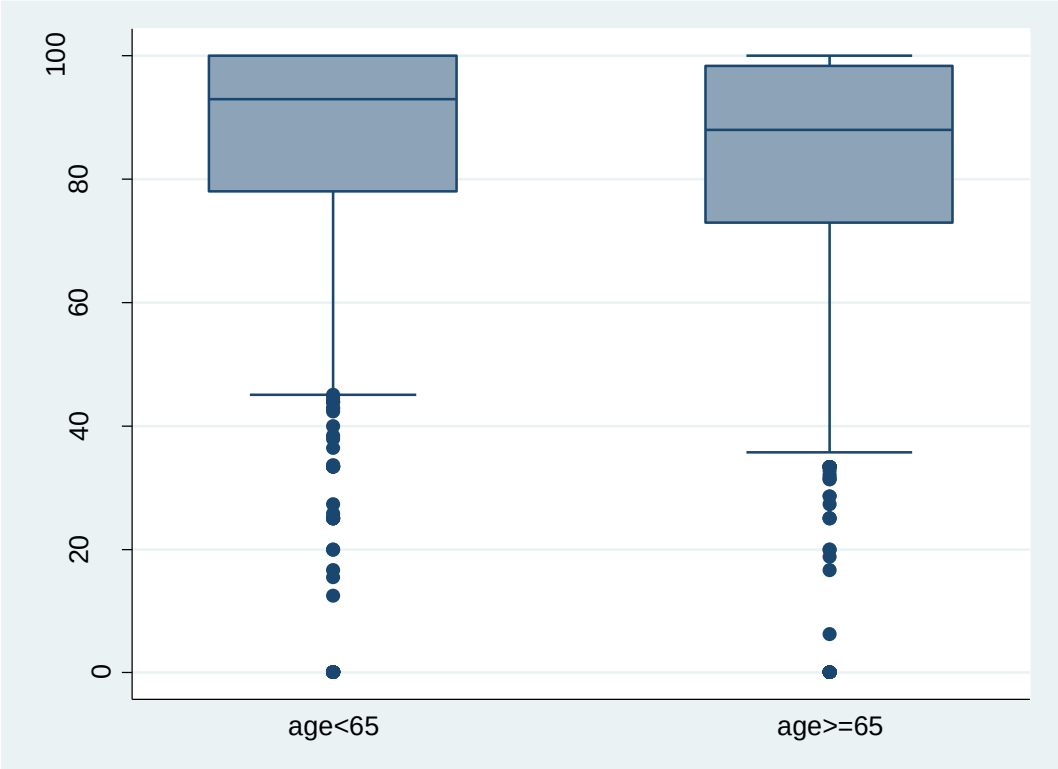



**Table 14. Distribution of the Performance of the ICD Discharge Composite Measure Stratified by Age at the Hospital-Level 2018 (N=1,732)**

| Description | Age < 65 | Age ≥ 65 |
|---|---|---|
| Mean | 85.22% | 81.78% |
| Std Deviation | 20.87% | 20.91% |
| 100% Max | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% |
| 75% Q3 | 100.00% | 98.33% |
| 50% Median | 92.98% | 88.00% |
| 25% Q1 | 78.00% | 72.92% |
| 10% | 60.00% | 55.17% |
| 5% | 50.00% | 42.86% |
| 1% | 0.00% | 0.00% |
| 0% Min | 0.00% | 0.00% |

Median Scores test p < .0001

**Figure 12. Distribution of Performance of the Composite Measure at Discharge Stratified by Age Group at the Hospital-Level 2018**




***Race/Ethnicity***

The distribution of hospital performance was examined among White (non-Hispanic), Hispanic, Black (non-Hispanic), and Other race patients. In 2017, the median performance of the ICD Discharge Composite Measure among Hispanic, White (non-Hispanic), Black (non-Hispanic) and Other races is 100%, 87.9% 96.7%, and 100% respectively. In 2018, median performance among Hispanic, White (non-Hispanic), Black (non-Hispanic) and Other races is 100%, 88.8%, 100% and 100% respectively.

Interquartile ranges for each race/ethnicity are highlighted in Table 15 and Table 16 below.

The P-values indicate that the median performance of the composite among the different races is statistically significant. The results suggest that White non-Hispanic and Black non-Hispanic patients do not receive ICD medications upon discharge as frequently as Hispanic or Other race patients, however there are notably fewer Hispanic and Other race patients (collectively approx. 10% of cohort) in the dataset potentially contributing to noise due to power issues in those subgroups.

**Table 15. Distribution of Performance for the ICD Discharge Composite Measure Stratified by Race at the Hospital-Level 2017 (N= 1,732)**

| Description | Race | | | |
|---|---|---|---|---|
| | Hispanic | White non-Hispanic | Black non-Hispanic | Other |
| Mean | 84.06% | 82.80% | 84.41% | 86.46% |
| Std Deviation | 26.94% | 18.37% | 23.59% | 27.07% |
| 100% Max | 100.00% | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% | 100.00% | 100.00% |
| 75% Q3 | 100.00% | 97.37% | 100.00% | 100.00% |
| 50% Median | 100.00% | 87.87% | 96.67% | 100.00% |
| 25% Q1 | 76.47% | 73.68% | 76.47% | 85.71% |
| 10% | 50.00% | 60.00% | 50.00% | 50.00% |
| 5% | 0.00% | 50.00% | 33.33% | 0.00% |
| 1% | 0.00% | 0.00% | 0.00% | 0.00% |
| 0% Min | 0.00% | 0.00% | 0.00% | 0.00% |
| Median Scores Test | P<.0001 | | | |

**Figure 13. Distribution of the ICD Discharge Composite Measure Stratified by Race at the Hospital-Level 2017**

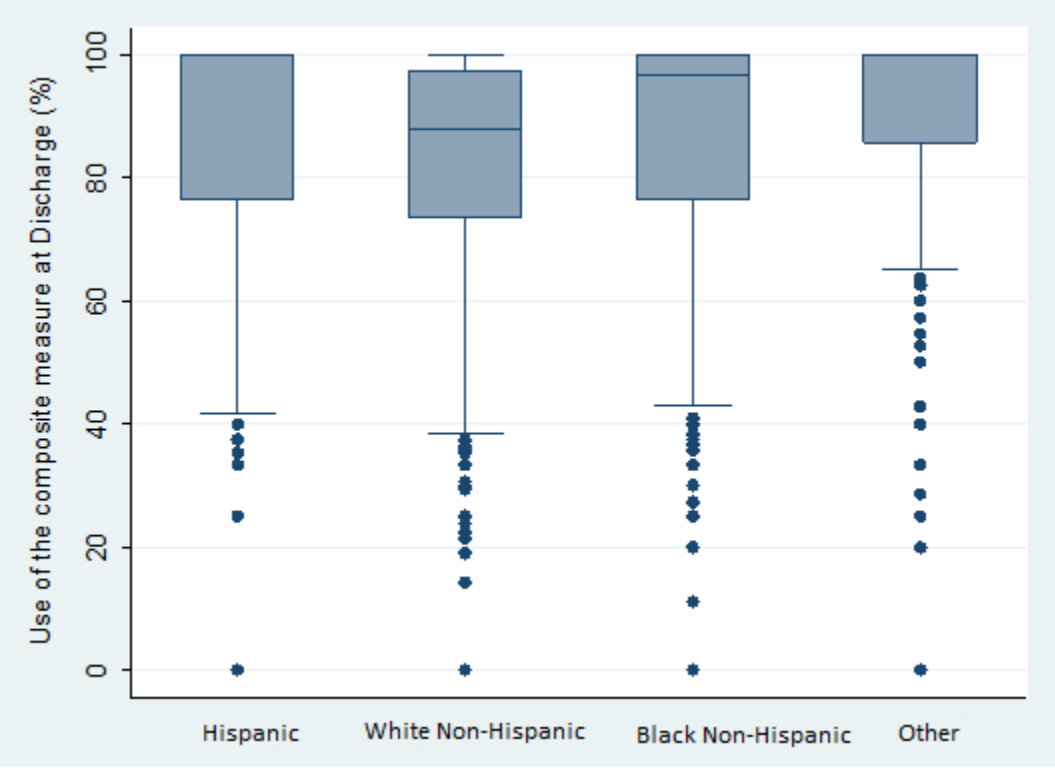
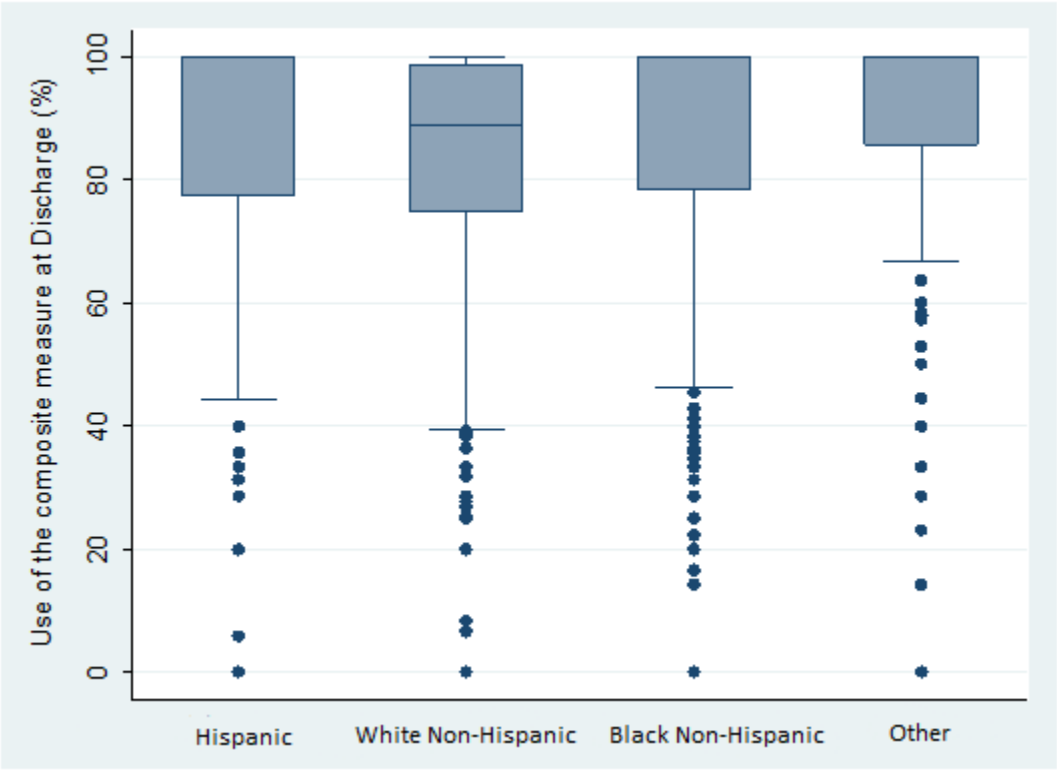



**Table 16. Distribution of Performance for the ICD Discharge Composite Measure Stratified by Race at the Hospital-Level 2018 (N= 1,732)**

| Description | Race | | | |
|---|---|---|---|---|
| | Hispanic | White non-Hispanic | Black non-Hispanic | Other |
| Mean | 84.76% | 82.68% | 84.84% | 86.02% |
| Std Deviation | 26.16% | 20.47% | 24.60% | 28.28% |
| 100% Max | 100.00% | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% | 100.00% | 100.00% |
| 75% Q3 | 100.00% | 98.53% | 100.00% | 100.00% |
| 50% Median | 100.00% | 88.75% | 100.00% | 100.00% |
| 25% Q1 | 77.42% | 74.83% | 78.42% | 85.71% |
| 10% | 50.00% | 58.33% | 50.00% | 50.00% |
| 5% | 0.00% | 43.75% | 33.33% | 0.00% |
| 1% | 0.00% | 0.00% | 0.00% | 0.00% |
| 0% Min | 0.00% | 0.00% | 0.00% | 0.00% |
| Median Scores Test | P<.0001 | | | |

**Figure 14. Distribution of the ICD Discharge Composite Measure Stratified by Race at the Hospital-Level 2018**




**Proportion of Dual Eligible Medicare and Medicaid Patients**

Hospitals (N=1,732) were stratified into quartiles by the proportion of dual Medicare and Medicaid patients. In 2017, the median performance for those hospitals with the fewest dual patients (Q1) was 84.5% (IQR: 68.2% to 100%). Among those hospitals with the highest proportion of dual patients (Q4), the median performance was 86.7% (IQR: 73.8% to 95.7%) (Table 17, Figure 15).

In 2018, median performance for hospitals with the fewest dual patients (Q1) was 83.3% (IQR: 66.7% to 100%). Among those hospitals with the highest proportion of dual patients (Q4), the median performance was 86.4% (IQR: 75% to 96.2%) (Table 18, Figure 16). Overall, hospitals with fewer dual eligible patients perform worse those with a higher proportion dual eligible patients.

**Table 17. Distribution of Performance Rate for the ICD Discharge Composite Measure Stratified by Dual Eligible Patients at the Hospital-Level 2017 (N=1,732)**

| Description | Dual Eligible (%) | | | |
|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 |
| Mean | 78.31% | 85.67% | 86.12% | 82.68% |
| Std Deviation | 23.43% | 13.22% | 13.12% | 16.24% |
| 100% Max | 100.00% | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 98.92% | 99.39% | 100.00% |
| 75% Q3 | 100.00% | 96.72% | 96.84% | 95.74% |
| 50% Median | 84.51% | 89.21% | 90.00% | 86.67% |
| 25% Q1 | 68.18% | 77.44% | 77.89% | 73.77% |
| 10% | 45.76% | 66.67% | 68.66% | 61.54% |
| 5% | 33.33% | 60.56% | 62.32% | 50.00% |
| 1% | 0.00% | 48.39% | 45.45% | 30.00% |
| 0% Min | 0.00% | 12.50% | 22.73% | 0.00% |

**Figure 15. Distribution of ICD Discharge Composite Measure Stratified by by Dual Eligible Patients at the Hospital-Level 2017**

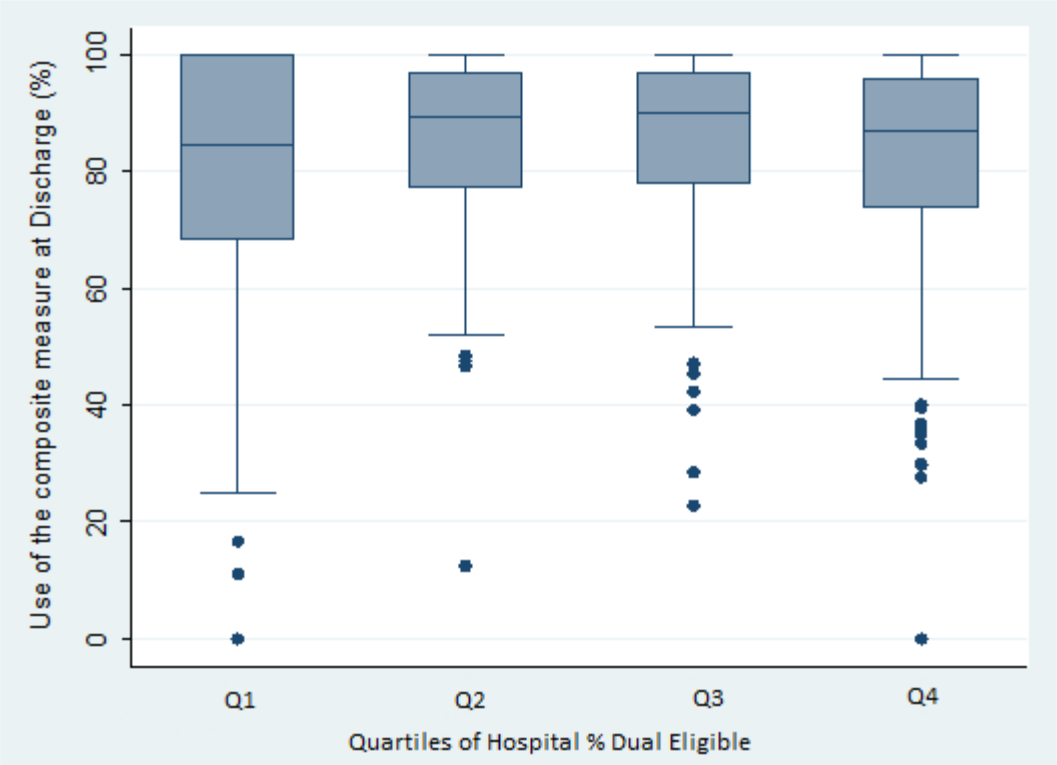
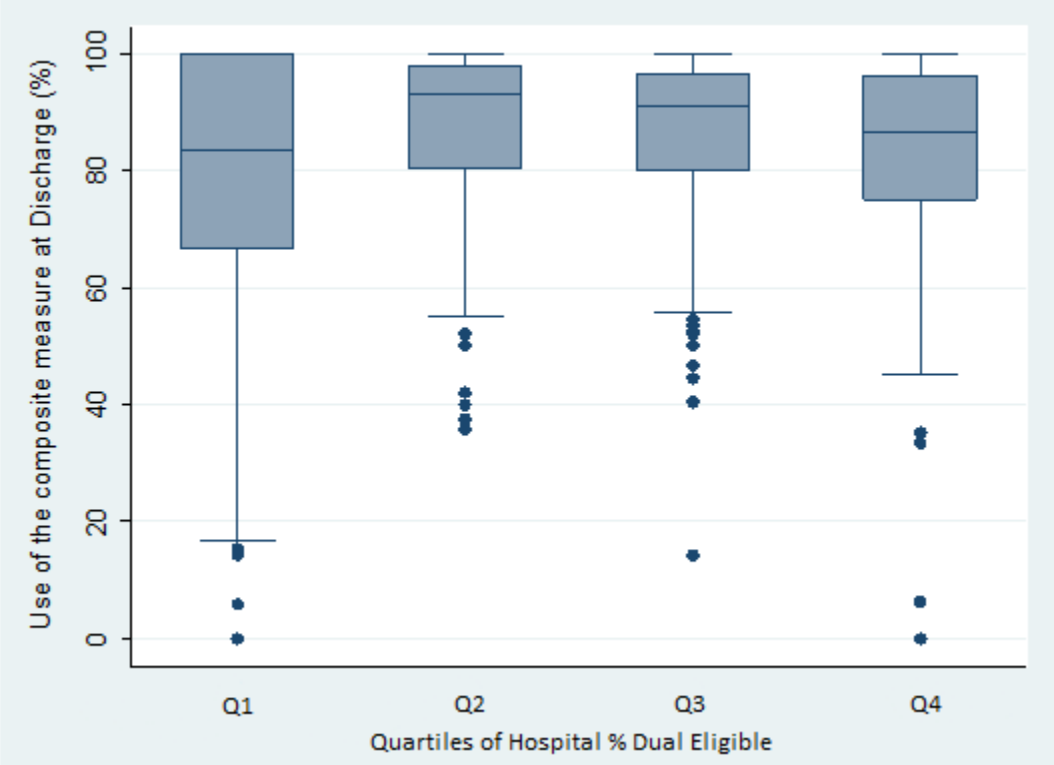



**Table 18. Distribution of Performance Rate for the ICD Discharge Composite Measure Stratified by Dual Eligible Patients at the Hospital-Level 2018 (N=1,732)**

| Description | Dual Eligible (%) | | | |
|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 |
| Mean | 76.28% | 88.24% | 86.65% | 83.04% |
| Std Deviation | 26.67% | 12.49% | 13.01% | 15.94% |
| 100% Max | 100.00% | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% | 99.56% | 100.00% |
| 75% Q3 | 100.00% | 97.73% | 96.55% | 96.15% |
| 50% Median | 83.33% | 92.86% | 90.91% | 86.41% |
| 25% Q1 | 66.67% | 80.47% | 80.00% | 75.00% |

| Description | Dual Eligible (%) | | | |
|---|---|---|---|---|
| | Q1 | Q2 | Q3 | Q4 |
| 10% | 40.00% | 72.07% | 68.02% | 61.54% |
| 5% | 5.88% | 65.63% | 61.76% | 53.13% |
| 1% | 0.00% | 40.00% | 46.67% | 33.33% |
| 0% Min | 0.00% | 35.71% | 14.29% | 0.00% |

**Figure 16. Distribution of ICD Discharge Composite Measure Stratified by by Dual Eligible Patients at the Hospital-Level 2018**



Across stratified analyses based on sex, age, race, and proportion of patients who are insured through Medicaid, we found significant overlap in the distribution of hospital performance.

**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?*)

2019 submission

The gap in performance rates, along with broad interquartile ranges, across various stratified populations demonstrates that this measure is necessary to improve the quality gap.

2015 submission

Given the gaps in care, there continues to be an opportunity for improvement.

---

**2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS**

***Note:*** *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*

***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*)

Not applicable

**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*)

Not applicable

**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?*)

Not applicable

---

**2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

***Note:*** *Applies to the overall composite measure.*

**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*)

2019 submission

No additional testing was completed.

2015 submission

The composite discharge medication measure is specified such that cases with missing data are assumed to have not met the metric. The performance ranges throughout this application reflect this approach. By following this method, the scores should be a true depiction of performance scores.

Missing data defaults to "performance not met". This measure assumes that missing documentation on the process results in a failure of meeting a evidence based therapy.

**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*)

2019 submission

No additional testing was completed.

2015 submission

As noted above, there are no "discretionary" exclusions. All exclusions are necessary to the accurate calculation of performance of the measure. See section 2b3.2.

**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*)

2019 submission

No additional testing was completed.

2015 submission

No empirical analysis was performed. However, it was felt that the method employed would minimize the potential for gaming.

Given the low frequency of exclusions, we do not believe that the exclusions have any impact on the validity, accuracy or interpretability of this measure. The exclusions have little potential for bias especially given the ICD Data Quality Program audits all essentially performance measure elements on a 3 year cycle and would detect misclassifications of patient records.

**2c. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH**

***Note:*** *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

**2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.**

2019 submission

The empirical validity analysis demonstrated that the individual component measures fit the overall quality construct by assessing the Pearson correlation of the discharge medications composite measure with its components, including: ACE/ARB and Beta Blockers.

2015 submission

We believe the content validity of this measure has been achieved by virtue of the noted expertise of those individuals who developed this measure. The individual components of the composite have already shown to impact clinical outcomes. However the empirical analysis demonstrating the individual component measures fit the overall quality construct is currently being researched. The testing will focus on construct validation which will test the hypothesis on the theory of the construct that following these processes for patients with ICD implantations lead to better outcomes. This research is expected to ultimately be published in the medical literature.

**2d1.1 Describe the method used** (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2019 submission

We computed hospital-level measures for the two measure components individually and then correlated the results with the hospital-level composite results using Pearson correlation.

Additionally we conducted a logistic regression analysis at the hospital-level to examine the overall contribution of the individual component measures to the variance explained by the overall composite measure.

**2d1.2. What were the statistical results obtained from the analysis of the components?** (e.g., *correlations, contribution of each component to the composite score, etc*.; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

2019 submission

In 2017, the Pearson correlation coefficients between the discharge composite medication measure and its components were: ACE/ARB (r= 0.9185) and Beta Blockers (r= 0.7026). In 2018, Person correlation coefficients were: ACE/ARB (r= 0.9133) and Beta Blockers (r=0.7089).

**Table 19. Distribution of Performance of the ICD Discharge Composite Measure and its Components 2017 (N=1,732)**

| Description | ICD Discharge Composite Measure | ACE/ARB | Beta Blockers |
|---|---|---|---|
| Mean | 82.72% | 83.69% | 93.74% |
| Std Deviation | 18.00% | 17.54% | 10.32% |
| | | | |
| 100% Max | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% | 100.00% |
| 75% Q3 | 96.84% | 97.22% | 100.00% |
| 50% Median | 87.56% | 88.24% | 96.88% |
| 25% Q1 | 74.29% | 75.00% | 91.67% |
| 10% | 60.00% | 61.90% | 85.19% |
| 5% | 48.39% | 50.00% | 77.78% |
| 1% | 0.00% | 0.00% | 50.00% |
| 0% Min | 0.00% | 0.00% | 0.00% |

**Table 20. Distribution of Performance of the ICD Discharge Composite Measure and its Components 2018 (N=1,732)**

| Description | ICD Discharge Composite Measure | ACE/ARB | Beta Blockers |
|---|---|---|---|
| Mean | 82.54% | 83.94% | 93.34% |
| Std Deviation | 19.80% | 18.68% | 12.45% |
| 100% Max | 100.00% | 100.00% | 100.00% |
| 99% | 100.00% | 100.00% | 100.00% |
| 95% | 100.00% | 100.00% | 100.00% |
| 90% | 100.00% | 100.00% | 100.00% |
| 75% Q3 | 97.41% | 97.92% | 100.00% |
| 50% Median | 87.95% | 89.32% | 97.20% |
| 25% Q1 | 75.00% | 76.15% | 91.67% |
| 10% | 59.80% | 62.50% | 84.00% |
| 5% | 46.67% | 50.00% | 76.19% |
| 1% | 0.00% | 0.00% | 25.00% |
| 0% Min | 0.00% | 0.00% | 0.00% |

**Table 21. Logistic Regression Model of ICD Discharge Composite Measure and its Components 2017**

| | Logistic Regression Model | |
|---|---|---|
| | C | Variance |
| Overall | | 0.11444 |
| Explained by | | |
| ACE/ARB | 0.891 | 0.08653 |
| Beta Blocker | 0.677 | 0.03679 |
| Both | 1.000 | 0.11443 |

**Table 22. Logistic Regression Model of ICD Discharge Composite Measure and its Components 2018**

| | Logistic Regression Model | |
|---|---|---|
| | C | Variance |
| Overall | | 0.11444 |
| Explained by | | |
| ACE/ARB | 0.891 | 0.08653 |
| Beta Blocker | 0.677 | 0.03679 |

| | | |
|---|---|---|
| Both | 1.000 | 0.11443 |

In addition, a logistic regression analysis was performed to examine the overall contribution of each of the individual component measures to the variance explained by the overall composite measure. ACE/ARB and Beta Blockers explained 89.0% and 68.0% of the overall variance, respectively.

**2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite?** (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

2019 submission

A correlation coefficient of 0.6 or higher is considered a 'strong correlation'. The results of the empirical validity testing demonstrate a strong correlation between the discharge medication composite and all of its components, meaning both ACE/ARB and Beta Blockers contribute individually to the overall composite measure that included both types of discharge medications. These results also suggest the components of the measure significantly explain variance in performance and prediction.

Reference:

Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal, 24*(3), 69-71.

**2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible**

**2d2.1 Describe the method used** (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2019 submission

This is an all-or-none composite; thus, no empirical analyses pertinent to aggregations or weighting were conducted. The components mentioned throughout the application are part of the composite measure indicator definition, not the composite of different measures.

**2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules?** (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

2019 submission

This all-or-none composite method indicates that each of the individual measure components were weighed equally.

**2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct?** (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

2019 submission

This all-or-none composite has each of the individual measure components weighed equally based on the strong clinical recommendations and studies demonstrating that patients who are prescribed each of these medications will have better outcomes such as reduced readmission and mortality rate at six months. As a result, it would not be appropriate to apply different weighting where compliance with one component influences a facility's performance score more than the other.

**Distribution of frequency of use of the composite measure and its components**

Pursuant to the request of the NQF, we have provided a table which presents the distribution of the use of the composite measure and of its medication components, ACEI/ARBs and Beta Blockers. Information on this distribution of this is shown below in table 1. Information on the distribution of the performance of the composite measure at the hospital level is detailed further in section 1d.2 of the evidence supplement.

**Table 1.**

**Distribution of The Composite Measure and Its Components**

| Description | Composite Measure | | ACEI/ARB | | Beta Blocker | |
|---|---|---|---|---|---|---|
| | Volume | Value | Volume | Value | Volume | Value |
| N | 1606 | 1606 | 1596 | 1596 | 1606 | 1606 |
| Mean | 121.77 | 0.7790 | 99.2180 | 0.8105 | 120.5479 | 0.9135 |
| Std Deviation | 133.13 | 0.1653 | 105.2580 | 0.1481 | 131.7408 | 0.1111 |
| 100% Max | 1062 | 1.0000 | 906.0000 | 1.0000 | 1061.0000 | 1.0000 |
| 99% | 577 | 1.0000 | 477.0000 | 1.0000 | 575.0000 | 1.0000 |
| 95% | 388 | 1.0000 | 315.0000 | 1.0000 | 380.0000 | 1.0000 |
| 90% | 296 | 0.9728 | 239.0000 | 0.9880 | 292.0000 | 1.0000 |
| 75% Q3 | 168 | 0.8913 | 134.0000 | 0.9124 | 166.0000 | 0.9750 |
| 50% Median | 83 | 0.7928 | 69.0000 | 0.8227 | 82.0000 | 0.9350 |
| 25% Q1 | 27 | 0.7059 | 24.0000 | 0.7436 | 26.0000 | 0.8889 |
| 10% | 8 | 0.5938 | 7.0000 | 0.6538 | 8.0000 | 0.8333 |
| 5% | 4 | 0.5000 | 3.0000 | 0.5600 | 4.0000 | 0.7671 |
| 1% | 1 | 0.0000 | 1.0000 | 0.2329 | 1.0000 | 0.4286 |
| 0% Min | 1 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| **Correlation coefficient between DCM and Others | | | 0.8709 | | 0.7255 | |

## 3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

**3a. Byproduct of Care Processes**

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

**3a.1. Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, 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 clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

**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).

There is no eCQM specification for this measure.

**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.**

Availability:

Participating hospitals report patient demographics, medical history, risk factors, hospital presentation, initial cardiac status, procedural details, medications, laboratory values and in-hospital complications. All of the data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient

population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions can manually report using a free web-based tool or automate the reporting by using certified software developed by third-party vendors. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

Sampling:

There is no sampling of patient data allowed within the contractual terms of participation in the ICD Registry in NCDR. Section 2.b of the NCDR Master Agreement with participants includes 'Participant Responsibilities': "b. Use of ACCF Data Set and ACCF-Approved Software. Participant will submit a data record on each patient who receives medical care and who is eligible for inclusion in the Registries in which Participant is participating under this Agreement." Adult patients, ages 18 years and older, who have an ICD implanted. Patients are selected for inclusion by reviewing existing medical records and no direct interaction with the patient will be required outside of the normal course of care. There will be no discrimination or bias with respect to inclusion on the basis of sex, race, or religion.

Patient confidentiality:

Patient confidentiality is preserved as the data are in aggregate form. The ICD Registry dataset, comprised of approximately 320 data elements, was created by a panel of experts using available ACC-AHA guidelines, data elements and definitions, and other evidentiary sources. Private health information (PHI), such as social security number, is collected. The intent for collection of PHI is to allow for registry interoperability and the potential for future generation of patient-level drill downs in Quality and Outcomes Reports. Registry sites can opt out of transmitting direct identifiers to the NCDR, however, so inclusion of direct identifiers in the registry is at the discretion of the registry participants themselves. When using the NCDR web-based data collection tool, direct identifiers are entered but a partition between the data collection process and the data warehouse maintains the direct identifiers separate from the analysis datasets. The minimum level of PHI transmitted to the ACCF when a participant opts out of submitting direct identifiers meets the definition of a Limited Dataset as such term is defined by the Health Insurance Portability and Accountability Act of 1996.

Data collection within the NCDR conforms to laws regarding protected health information. Patient confidentiality is of utmost concern. The proposed measure does not include a patient survey. Physician and/or institutional confidentiality are maintained by de-identified dashboard reports. There is no added procedural risk to patients through involvement in the ICD Registry. No testing, time, risk, or procedures beyond those required for routine care will be imposed. The primary risk associated with this measure is the potential for a breach of patient confidentiality. The ACCF has established a robust plan for ensuring appropriate and commercially reasonable physical, technical, and administrative safeguards are in place to mitigate such risks.

Data are maintained on secure servers with appropriate safeguards in place. The project team periodically reviews all activities involving protected health information to ensure that such safeguards including standard operating procedures are being followed. The procedure for notifying the ACCF of any breach of confidentiality and immediate mitigation standards that need to be followed is communicated to participants. ACCF limits access to Protected Health Information, and to equipment, systems, and networks that contain, transmit, process or store Protected Health Information, to employees who need to access the PHI for purposes of performing ACCF's obligations to participants who are in a contractual relationship with the ACCF. All PHI are stored in a secure facility or secure area within ACCF's facilities which has separate physical controls to limit access, such as locks or physical tokens.

The secured areas are monitored 24 hours per day, 7 days per week, either by employees or agents of ACCF by video surveillance, or by intrusion detection systems.

Each participant who has access to the NCDR website must have a unique identifier. The password protected webpages have implement inactivity time-outs. Encryption of wireless network data transmission and

authentication of wireless devices containing NCDR Participant's information ACCF's network is required. Protected Health Information may only be transmitted off of ACCF's premises to approved parties, which shall mean: A subcontractor who has agreed to be bound by the terms of the Business Associate Agreement between the ACCF and the NCDR Participant.

Time of Data collection:

1 Full time employee can enter on average roughly 1200 patient records per year (citation: ACC Marketing Intelligence Team).

**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)*.

The ACCF's program the National Cardiovascular Data Registry (NCDR) provides evidence-based solutions for cardiologists and other medical professionals committed to excellence in cardiovascular care. NCDR hospital participants receive confidential benchmark reports that include access to measure macro specifications and micro specifications, the eligible patient population, exclusions, and model variables (when applicable). In addition to hospital sites, NCDR Analytic and Reporting Services provides consenting hospitals' aggregated data reports to interested federal and state regulatory agencies, multi-system provider groups, third-party payers, and other organizations that have an identified quality improvement initiative that supports NCDR-participating facilities. Lastly, the ACCF also allows for licensing of the measure specifications outside of the Registry.

It should be noted that the centers already have to participate in this specific registry for reimbursement purposes so that currently almost all hospitals that implant ICDs in Medicare populations already participate. Hence there is no additional cost.

Measures that are aggregated by ACCF and submitted to NQF are intended for public reporting and therefore there is no charge for a standard export package. However, on a case by case basis, requests for modifications to the standard export package will be available for a separate charge.

## 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) |
|---|---|

| | |
|---|---|
| Not in use | Public Reporting<br>NCDR Public Reporting<br>https://cvquality.acc.org/ncdr-home/acc-public-reporting<br>NCDR ICD Registry™<br>https://cvquality.acc.org/NCDR-Home/registries/hospital-registries/icd-registry<br>Quality Improvement (external benchmarking to organizations)<br>NCDR National Outcomes Report<br>http://cvquality.acc.org/login |

**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

Voluntary Hospital Public Reporting Program: Hospitals may opt to publicly report their measure results based on data from the National Cardiovascular Data Registry (NCDR). Hospitals that choose to participate have their results displayed on ACC's CardioSmart. Currently Hospitals can report on the following NQF-endorsed measures:
NQF #0965: Use of all recommended medications (ACEI or ARB and beta-blocker) to improve heart function and blood pressure after ICD implant.
NQF # 0964: Therapy with aspirin, P2Y12 inhibitor, and statin at discharge following PCI in eligible patients (composite measure)
NQF# 2377: Overall Defect Free Care Composite (identified on website as "Complete Heart Attack Care")
NCDR ICD Registry:
National quality improvement registry intended to improve the quality of care provided to patients receiving ICD therapy since its inception in 2005. It provides a streamlined, consolidated method of collecting, monitoring and reporting clinically relevant cardiovascular data within a framework that ensures both hospital and patient confidentiality. This enables participants to better focus on ACC/AHA guideline-recommended care and to develop new ways for the registry to advance improvements in care and examine newer clinical questions. There are over 1,600 participating sites with 1,449,976 cumulative records as of Q2 2019.

**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.**

Performance results are distributed to all ICD registry participants as part of quarterly benchmark reports, which provide a detailed analysis of an institution´s individual performance in comparison to the entire registry population from participating hospitals across the nation. Reports include an executive summary dashboard,

at-a-glance assessments, and patient level drill-downs. Registry participants also have access to an outcome report companion guide, which provides common definitions and detailed metric specifications to assist with interpretation of performance rates.

**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.**

Results are provided as part of quarterly performance report, which includes a rolling 4 quarters of data.

Participating hospitals in the ICD registry report on the following: patient demographics; provider and facility characteristics; adverse event rates; ICD performance measures and select quality measures and outcomes and compliance with ACC/AHA clinical guideline recommendations.

The majority of the required data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions can manually report using a free web-based tool or automate the reporting by using certified software developed by third-party vendors. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

There are a number of methods used to educate and provide general support to registry participants. This includes the following:

- Registry Site Manager Calls are available for all NCDR participants. RSM calls are provided as a source of communication between NCDR and participants to provide a live chat Q and A session on a continuous basis.
- New User Calls are available for NCDR participants, and are intended for assisting new users with their questions.
- NCDR Annual Conference

The NCDR Annual Conference is a well-attended and energetic two-day program at which participants from across the country come together to hear about new NCDR and registry-specific updates. During informative general sessions, attendees can learn about topics such as transcatheter therapies, the NCDR dashboard, risk models, data quality and validation, and value-based purchasing. Attendees also receive registry updates and participate in advanced case studies covering such topics as Appropriate Use Criteria and outcomes report interpretation.

- Release notes (for outcomes reports)
- Clinical Support

The NCDR Product Support and Clinical Quality Consultant Teams are available to assist participating sites with questions Monday through Friday, 9:00 a.m. - 5:00 p.m. ET.

**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.**

Feedback is typically obtained through monthly registry site manager monthly calls, ad hoc phone calls tracked with salesforce software, and during registry –specific break-out sessions at the NCDR's annual meeting. Registry Steering Committee members may also provide feedback during regularly scheduled calls.

**4a2.2.2. Summarize the feedback obtained from those being measured.**

The data elements are clear and are supported by the guidelines.

The benefits to these measures are they are supported by the guidelines and promote process improvement initiatives. The ICD registry has stringent coding requirements for the medications and thus improved documentation is required to support the coding of these measures.

**4a2.2.3. Summarize the feedback obtained from other users**

The users reported that the ICD Registry helped them with the documentation at their sites and that enabled them to easily do quality improvement projects.

**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.**

The measure language was updated to further simplify and clarify the measure intent but the changes were not substantive. Specifically, the denominator was expanded to also include cardiac resynchronization therapy defibrillator (CRT-D) implant patients and the exclusions for the measure were removed since they were duplicative to what is captured and calculated in the numerator. These changes did not generate any feedback from participants.

**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.**

While the mean rate of performance for this composite across participating facilities was greater than 80% for 2017 and 2018, opportunities for improvement across facilities continue to exist with some facilities demonstrating low performance scores (<50% in the 5th percentile). Progress toward improvements overall has been made when the current mean rate is compared to the mean rate of 74% when the measure was first released (2011-2012).

**4b2. Unintended Consequences**

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

**4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**

There were no unintended consequences to individuals or populations identified during testing or implementation.

**4b2.2. Please explain any unexpected benefits from implementation of this measure.**

Sites have reported being able to develop process improvement mechanisms and improve their documentation practices as a result of implementing this measure.

## 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)**

0066 : Coronary Artery Disease (CAD): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy - Diabetes or Left Ventricular Systolic Dysfunction (LVEF < 40%)

0070 : Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)

0070e : Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)

0071 : Persistence of Beta-Blocker Treatment After a Heart Attack

0081 : Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)

0081e : Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) or Angiotensin Receptor-Neprilysin Inhibitor (ARNI) Therapy for Left Ventricular Systolic Dysfunction (LVSD)

0083 : Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)

0117 : Beta Blockade at Discharge

0236 : Coronary Artery Bypass Graft (CABG): Preoperative Beta-Blocker in Patients with Isolated CABG Surgery

0594 : Post MI: ACE inhibitor or ARB therapy

**5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.**

Note also 0696: STS composite score. section 5.1a. Confirmed with the NQF Quality Positioning System that this measure is still endorsed.

**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.**

Measure #0965 is a subset of other measures and the measures are completely harmonized with the exception of one area. It appears that only one measure (#81e) currently includes prescribing of ARNI as an acceptable therapy in the numerator. We assume that the other measures be updated to reflect the current evidence and there is no need for further harmonization.

**5b. Competing Measures**

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

**OR**

Multiple measures are justified.

**5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

**Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)**
N/A

## Appendix

**A.1 Supplemental materials may be provided in an appendix.** All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment Attachment: icd_v2_codersdatadictionary_2-2-637001858309276129-637061353942435374-637088191502603381.pdf

## Contact Information

**Co.1 Measure Steward (Intellectual Property Owner):** American College of Cardiology

**Co.2 Point of Contact:** Jarrott, Mayfield, Jmayfield@acc.org, 202-375-6572-

**Co.3 Measure Developer if different from Measure Steward:** American College of Cardiology

**Co.4 Point of Contact:** Beth, Denton, bdenton@acc.org, 202-375-6631-

## 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.**

For this particular topic those individuals who were involved in identifying the key attributes and variables for this process measure were leaders and experts in the field of electrophysiology. Serial phone calls were held to both define the eligible population and given process. These clinical leaders are noted below.

NCDR Clinical Subworkgroup ensured the measure demonstrated an opportunity for improvement, had strong clinical evidence, and was a reliable and valid measure. These members included Drs. Jeptha Curtis (Chair), Frederick Masoudi, John Rumsfeld, Matt Reynolds, and Mark Kremers.

NCDR Scientific Quality and Oversight Committee—a committee that served as the primary resource for crosscutting scientific and quality of care methodological issues. These members included Drs. Frederick Masoudi (Chair) , David Malenka, Thomas Tsai, Matthew Reynolds, David Shahian, John Windle, Fred Resnic, John Moore, Deepak Bhatt, James Tcheng, Jeptha Curtis, Paul Chan, Matthew Roe, and John Rumsfeld.

**Measure Developer/Steward Updates and Ongoing Maintenance**

**Ad.2 Year the measure was first released:** 2011

**Ad.3 Month and Year of most recent revision:** 02, 2015

**Ad.4 What is your frequency for review/update of this measure?** With dataset revisions and based on new evidence.

**Ad.5 When is the next scheduled review/update for this measure?** 11, 2019

**Ad.6 Copyright statement:** American College of Cardiology Foundation All Rights Reserved

**Ad.7 Disclaimers:** ACC realizes the various NCDR endorsed measures are not readily available on their own main webpage. However, ACCF plans to update their main webpage (acc.org) to include the macrospecifications of the NQF endorsed measures. ACC hopes to work collaboratively with NQF to create a consistent and standard format would be helpful for various end users. In the interim, the supplemental materials include the details needed to understand this model. In addition, interested parties are always able to contact comment@acc.org to reach individuals at the ACC Quality Measurement Team.

**Ad.8 Additional Information/Comments:** ACC appreciates the opportunity to submit measures for this NQF endorsement maintenance project.