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 #: 0083e

Corresponding Measures: 0083

De.2. Measure Title: Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)

Co.1.1. Measure Steward: PCPI Foundation

De.3. Brief Description of Measure: Percentage of patients aged 18 years and older with a diagnosis of heart failure (HF) with a current or prior left ventricular ejection fraction (LVEF) < 40% who were prescribed beta-blocker therapy either within a 12-month period when seen in the outpatient setting OR at each hospital discharge

1b.1. Developer Rationale: Beta-blockers are recommended for all patients with stable heart failure and left ventricular systolic dysfunction, unless contraindicated. Treatment should be initiated as soon as a patient is diagnosed with left ventricular systolic dysfunction and does not have low blood pressure, fluid overload, or recent treatment with an intravenous positive inotropic agent. Beta-blockers have been shown to lessen the symptoms of heart failure, improve the clinical status of patients, reduce future clinical deterioration, and decrease the risk of mortality and the combined risk of mortality and hospitalization.

Also, a 2011 analysis of IMPROVE HF data by Fonarow and colleagues revealed that all 4 current ACC/AHA HF outpatient performance measures were associated with decreased risk of 24-month mortality. For the 2 summary measures of HF care processes, there was also a strong positive association between greater conformity to the summary measures and improved risk-adjusted survival. These findings may have significant clinical and public health implications, providing evidence to suggest that current, and some emerging, outpatient process measures may effectively reflect the quality of care provided to patients with HF who are treated in outpatient practice settings.

S.4. Numerator Statement: Patients who were prescribed beta-blocker therapy either within a 12-month period when seen in the outpatient setting OR at each hospital discharge

S.6. Denominator Statement: All patients aged 18 years and older with a diagnosis of heart failure with a current or prior LVEF < 40%

S.8. Denominator Exclusions: Denominator Exceptions:

Documentation of medical reason(s) for not prescribing beta-blocker therapy (e.g., low blood pressure, fluid overload, asthma, patients recently treated with an intravenous positive inotropic agent, allergy, intolerance, other medical reasons).

Documentation of patient reason(s) for not prescribing beta-blocker therapy (e.g., patient declined, other patient reasons).

Documentation of system reason(s) for not prescribing beta-blocker therapy (e.g., other reasons attributable to the healthcare system).

De.1. Measure Type: Process

S.17. Data Source: Electronic Health Records

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

IF Endorsement Maintenance – Original Endorsement Date: Mar 14, 2016 **Most Recent Endorsement Date:** Mar 14, 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**

Evidance Summary

- No changes in evidence since the 2016 evaluation. The developer continues to cite the 2013 ACCF/AHA guideline for the management of heart failure (Class 1, Level A). The developers note that while there have been focused updates on the guidelines in 2014, the recommendations remain unchanged.

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?*

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency: high (Box 5) → High (Box 5a) → High

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 provided the following EHR performance data from CMS's PQRS program from January 2016 to December 2016.
    - Number of quality events: 52,213
    - Mean: 0.72
    - Standard Deviation: 0.32
    - Minimum: 0.00
    - Maximum: 1.00
    - Interquartile Range 0.50 (1.00-0.50)
- The performance data does not include the number of providers (measured entity) used to calculate the performance rates provided.
- The EHR average performance rate reported for the 2018 MIPS benchmark report is 73.2% and standard deviation of 19.6.
- The average performance rate reported for the 2018 MIPS benchmark report is 73.2% with a standard deviation of 19.6.
- The developer also provided a summary of data from the literature.

Disparities

- No new disparities information was provided. The developer noted that while this measure is included in federal programs, no disparities data have been made available to analyze and report.
    - Disparities data is required for maintenance of endorsement.
- The developer provided a summary of disparities data from the literature.

Questions for the Committee:

- Can a gap in gare be determined if the number of providers is not included in the performance data?
- If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?

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

RATIONALE: Unable to determine gap in care without number of providers included in the performance data.

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

1a. Evidence:

- No concerns
- Satisfactory
- The evidence is strong that bbs improve outcomes in LVSD

1b. Performance Gap:

- Significant. Though unclear to me if this reflects missing data/lack of feasibility. Also no data on disparities
- Data on the gap in care did not include the number of provoiders
- Based on the PQRS program, performance is only 0.72 with an IQR of 0.50. Disparities data and subgroup data are not available.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

eCQM Technical Advisor(s) review:

Submitted measure is an HQMF compliant eCQM	The submitted eCQMspecifications follow the industry accepted format for eCQM (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF,QDM, or CQL limitations	N/A – All components in the measure logic of the submitted eCQM are represented using the HQMF,QDM, or CQL standards

Value Sets	The submitted eCQM specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously. – this includes 100% coverage of measured patient population testing with pass/fail test cases for each population;

Feasibility Testing	Number of data elements included in measure calculation: 29 Number of data elements scoring less than 3 on scorecard: 23 Number of data elements not assessed on scorecard: 1 LeftVentricularSystolicDysfunction_Diagnosis • low scoring domains: accuracy, standards, workflow OutpatientConsultation_EncounterPerformed • low scoring domains: availability, accuracy, standards, workflow • Professional billing (CPT) not captured in EHR; Planning to implement Cerner billing Arrhythmia_Diagnosis • low scoring domains: availability, accuracy, workflow CareServicesinLong-TermResidentialFacility_EncounterPerformed • low scoring domains: availability, accuracy, standards, workflow • Services not provided by IU Health HomeHealthcareServices_EncounterPerformed • low scoring domains: availability, accuracy, standards, workflow • Services not provided by IU Health HeartFailure_Diagnosis • low scoring domains: workflow NursingFacilityVisit_EncounterPerformed • low scoring domains: availability, accuracy, standards, workflow • Services not provided by IU Health PatientProviderInteraction_EncounterPerformed • low scoring domains: availability, accuracy, standards, workflow • Professional billing (CPT) not captured in EHR ModerateorSevereLVSD_Diagnosis • low scoring domains: availability, accuracy, workflow OfficeVisit_EncounterPerformed

	• low scoring domains: availability, accuracy, standards, workflow • Professional billing (CPT) not captured in EHR; Planning to implement Cerner billing CardiacPacerinSitu_Diagnosis • low scoring domains: workflow CardiacPacer_DeviceApplied • low scoring domains: availability, accuracy, standards, workflow EjectionFraction_DiagnosticStudyPerformed • low scoring domains: availability, accuracy, standards, workflow • Data imported from another application; available in EMR, but not codified BetaBlockerTherapyIngredient_Allergy/Intolerance • low scoring domains: availability, workflow Hypotension_Diagnosis • low scoring domains: workflow Asthma_Diagnosis • low scoring domains: accuracy, workflow Bradycardia_Diagnosis • low scoring domains: accuracy, workflow IntolerancetoBetaBlockerTherapy_Diagnosis • low scoring domains: availability, accuracy, standards, workflow • Entered as allergy, not diagnosis BetaBlockerTherapyforLVSD_MedicationNotOrdered • low scoring domains: availability, workflow BetaBlockerTherapyforLVSD_MedicationOrder • low scoring domains: workflow DischargeServices-HospitalInpatient_EncounterPerformed • low scoring domains: availability, accuracy, standards, workflow

	• Professional billing (CPT) not captured in EHR; Planning to implement Cerner billing AllergytoBetaBlockerTherapy_Diagnosis • low scoring domains: availability, accuracy, standards, workflow • Entered as allergy, not diagnosis 8867-4_PhysicalExamPerformed • low scoring domains: availability, accuracy, standards, workflow How is the data element used in computation of measure? How the data element is feasibile within the context of the measure logic? What is the plan for readdressing the data element?

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

Evaluators: NQF Staff

Reliability Summary:

- The measure will be considered for endorsement at the clinician group level of analysis and outpatient setting only unless additional testing is provided.
- Per the 2013 NQF eCQM Feasibility Assessment Technical Report, the balance between feasibility and validity/reliability and the usefulness of a measure is critical. Data element validity and data accuracy often overlap. Data accuracy (feasibility) is intended to assess the likely "correctness" of a data element prior to formal reliability and validity testing.
- Feasibility testing results (see above and Criterion 3.) identified substantial feasibility issues with 23 out of 29 data elements, including the critical data elements (numerator, denominator, exceptions), though the feasibility assessment is different from reliability and validity testing.

Questions for the Committee regarding reliability:

- Does the Committee have any concerns about effect of the feasibility results on the reliability of the measure?
- Seek clarification from the developer to determine if the reliability scores are the average reliability for providers with 1+ events and 10+ events.
- Based on the results of the signal-to-noise analysis, scope of testing and measure specification concerns, is the Committee certain and/or confident that the performance measure scores are reliable?

Validity Summary:

- According to the 2013 NQF eCQM Feasibility Assessment Technical Report, quality data needs to fit into the clinical workflow in order to be recorded at the point of care by authoritative sources. It is of little benefit to have the capability of capturing certain patient symptoms if it requires five clicks and three screens during a busy clinical encounter, for the end result will likely be missing data."
- The feasibility assessment provided shows difficulty capturing the majority of the data elements during the course of care (workflow).
- The developer did not empirically assess the extent and distribution of missing data or nonresponse; this is required.

Questions for the Committee regarding validity:

- Does the Committee have any concerns about the effect of the feasibility results on the validity of the measure?
- Does the Committee have any concerns about the effect of potential missing data and/or nonresponse on the performance scores?
- Based on the results of the correlation analysis, scope of testing and analysis of potential threats, is the Committee certain and/or confident that the performance measure scores are a valid indicator of quality?

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

RATIONALE: Unable to determine the reliability of the measure due to the concerns identified about the measure specifications.

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

RATIONALE: Unable to determine the validity of the measure because potential threats relevant to the measure were not empirically assessed.

Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0083e

Measure Title: e-measure Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)

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.**
   - Levels of analysis and care settings inconsistent with testing provided. The level of analysis (LoA) specified are for individual clinicians and clinician groups. The care settings specified are home care, inpatient/hospital, other, outpatient, domiciliary, and nursing facility.. The LoA and care settings in the measure specifications must align with testing (clinician group and outpatient services).
   - Additional testing is required for endorsement at the individual clinician level in home care, inpatient/hospital, other, domiciliary, and nursing facility setting.
   - Section 1.5 and 1.6 discuss minimum number of quality reporting events (10) and providers who had 10 or more patients eligible for this measure.
     - The difference between reporting events and patients is not clear.
     - Minimum number of patients and/or reporting events is not included in specifications.

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**
   - Reliability testing conducted at clinician group level of analysis in outpatient setting only.

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
   - Reliablity testing conducted at the score level using signal to noise ratio.
   - Providers must have at least 10 eligible reporting events to be included in calculation – this is inconsistent with specifications.
   - Specifications include outpatient and inpatient settings (see above); developer did not provide testing for both outpatient setting and inpatient/hospital setting. NQF criteria states that testing must be conducted for the measure as specified.

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

   Submission document: Testing attachment, section 2a2.3
   - Reliability for 1+ events: 0.81; 10+ events: 0.86. Developer does not state if these results are the average reliability for providers.

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

Submission document: Testing attachment, section 2a2.2

☐ **Yes**

☒ **No**

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

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

Submission document: Testing attachment, section 2a2.2

☐ **Yes**

☐ **No**

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

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

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

☐ **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.**

- Unable to determine level of certainty or confidence that the performance measure scores are reliabile based on the reliability statistic and scope of testing due to the concerns about the measure specifications. Further clarification needed about outpatient and inpatient/hospital setting included in specifications.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

- Current testing data states providers with minimum (10) number of quality reporting events – this is inconsistent with specifications.
- Data demonstrates 3,168 exceptions reported and average number of exceptions per provider (3.37). NQF criteria for eCQMs states that if exclusions (or exceptions) are not based on the clinical evidence, analyses should identify the overall frequency of occurrence of the exclusions as well as variability across the measured entities to demonstrate the need to specify exclusions.

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

Submission document: Testing attachment, section 2b4.

- NQF criteria for eCQMs states this criterion is about using the measure as specified to distinguish differences in performance across the entities that are being measured. The performance measure scores should be computed for all accountable entities for which eCQM data are available (not just those on which

reliability/validity testing was conducted) and then analyzed to identify differences in performance. The developer only provided results based on sample used for reliability testing.

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.
    - Developer did not perform test to identify the extent and distribution of missing data or nonresponse. This is required because different uses of an EHR data field by clinicians or different data processing or extraction protocols in different EHRs can result in incorrect or missing data and produce different performance scores.
    - The feasibility assessment provided shows difficultly capturing the majority of the data elements during the course of care (workflow). The eCQM Feasibility Report states, "the end results will likely be missing data."

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**

VALIDITY: TESTING

17. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**
18. **Method of establishing validity of the measure score:**

    ☐ **Face validity**

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

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

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

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

- Correlation analysis was conducted for validity testing using the performance measure score on this measure (NQF #0083e) and another eCQM, NQF #0081e: *Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)* due to similarities in patient population and domain.
- The developers hypothesize a positive relationship between the two measures.

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

Submission document: Testing attachment, section 2b2.3

- Per developer, this measure has a strong positive correlation (0.65) with another evidence-based process of care measure (NQF #0081e).

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

Submission document: Testing attachment, section 2b1.

☐ **Yes**

☐ **No**

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

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

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

☐ **Yes**

☐ **No**

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

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

☐ **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.)

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

- Potential threats to validity that are relevant to the measure not empirically assessed; therefore, unable to determine validity of the measure.

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

2a1. Reliability-Specifications:

- Many if not most data elements appear to lack high availability/accuracy
- The value of the ejection fraction or a range option was not included in the data elements. Only documentation that the ejection fraction was measured is specified.
- The data seem to be clearly defined, but they are frequently not available

2a2. Reliability – Testing:

- Concerns about data availability/accuracy. Can this differ across practice or provider and create unreliable estimates as a result?
- Testing was only done at the outpatient setting but not at hospital discharge setting.
- Tested at the measure score only. Reliability testing conducted at clinician group level of analysis in outpatient setting only.

2b1. Validity -Testing:

- Am not sure that data testing reflected measure specifications about minimum number of quality reporting events
- Several data elements scored low in availability and accuracy. There is no actual data element tested for the ejection fraction percent.
- Validity was not tested at the level of patient data

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data):

- Concern per panel re: data collection in clinical workflow being very suboptimal. Also there does not appear to be testing done to describe amount of missing data or its impact on scores
- No missing data analysis was done, because submitted cases with missing data would be rejected from inclusionin the dataset. We don't have information on missing data
- Validity testing is incomplete

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment):

- Not risk adjusted.
- This is not risk adjusted. Exclusions averaged 3.37 per provider.

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.

- Developer provided feasibility testing results from two care settings: acute care; critical access; tertiary care; ambulatory services, etc. setting (IU) and clinician office/clinic ambulatory care setting (NUMC).
    - NUMC scorecard analysis (tab 4) shows measure is currently 51.0% feasible on a scale of 0% to 100%.
    - IU scorecard analysis (tab 4) shows measure is currently 80.0 % feasible on a scale of 0% to 100% - measure is not tested for inpatient setting/acute care setting.

- The developer indicated that all data elements are in defined fields in electronic health records (EHRs); however, feasibility assessment results provided show the following critical data elements are not currently available in a structured format within the EHRs tested:
    - Allergy to Beta Blocker Therapy
    - Allergy/Intolerance to Beta Blocker Therapy Ingredient
    - Arrhythmia
    - Beta Blocker Therapy for LVSD Medication Not Ordered
    - Cardiac Pacer Device Applied
    - Ejection Fraction Diagnostic Study Performed
    - Encounter Performed (LTSC, HH, and Nursing Facility)
    - Interolerance to Beta Blocker Therapy
    - Moderate or Severe LVSD
    - Patient Provider Interaction Encounter and various other Encounters
- The developer stated they have not identified any areas of concern as a result of testing and operational use and other feasibility issues.

Questions for the Committee:

- Is the Committee concerned that the eCQM is less feasible in the outpatient setting (as specified/tested) than in the inpatient/acute care setting (currently not tested in inpatient/acute care setting)?
- Do the eCQM Feasibility Scorecards demonstrate acceptable feasibility in multiple EHR systems and sites?
- Does the developer adequately address feasibility concerns?

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

RATIONALE: Feasibility scorecard indicates multiple critical data elements required to calculate the measure are not currently available in structured data fields.

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

- Concerns about about high missing data rates
- The actual percent ejection fraction is not specified. Several data elements were rated low on availability or accuracy.
- The feasibility assessment provided shows difficultly capturing the majority of the data elements during the course of 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

- This measure is currently used in the Merit-based Incentive Payment System (MIPS). The measure was previously used in the Physician Quality Reporting System (PQRS).
- The measure is not currently publicly reported, but data will be available for public reporting in Physician Compare beginning in late 2019.

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 no feedback received from those being measures resulted in any changes to the measure.

Additional Feedback:

- The developer also received feedback from the ONC Project Tracking System to clarify the difference between the two populations in the measure and explain the calculation of the single performance rate. The developer added the clarifying statements in 2018 and is in effect for 2019 reporting.

***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 improvement results were 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 states there have been no reports of unexpected findings from the implementation of this measure.

Potential harms

- The developer does not indicate any potential harms

Additional Feedback:

***Questions for the Committee**:*

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

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

RATIONALE: The developer did not discuss any progress on improvement.

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

4a1. Use - Accountability and Transparency:

- Used in MIPS and public reporting
- The measure is publicly reported in MIPS

4b1. Usability – Improvement:

- Concerns on feasibility may impact usability if there is high levels of missing information or erroneously entered information
- No trend over time data provided to see if care is improving
- The developer states there have been no reports of unexpected findings from the implementation of this measure. The developer does not indicate any potential harms

Criterion 5: Related and Competing Measures

Related or competing measures

- #2438 (endorsement removed): Beta-Blocker Therapy (i.e., bisoprolol, carvedilol, or sustained-release metoprolol succinate) for LVSD Prescribed at Discharge (The Joint Commission). The measure focus is the same but the level of analysis is different.
- #0070/e: 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
    - The developer states, "The specifications are harmonized to the extent possible. However, measure 0083 is focused on a patient population with heart failure and therefore the denominator

specifications for the measures (#0070/e and #0071) differ." Additionally, NQF 0071 is intended for use at the health plan level.

- 0083: Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)
  - This is the registry version of 0083e
- 0117: Beta Blockade at Discharge
  - The developer states "NQF 0117 is an inpatient/hospital level measure and includes only patients who have undergone isolated CABG surgery."
- 0127: Preoperative Beta Blockade
  - The developer states "NQF 0127 is also an inpatient/hospital level measure that focuses on administration of beta-blockers prior to isolated CABG surgery."

Harmonization

Developer states that these measures are harmonized to the extent possible.

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

- 0070e
- There are several related measures but none precisely compete

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

Developer Submission

NATIONAL QUALITY FORUM

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 0083e

Corresponding Measures: 0083

De.2. Measure Title: Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)

Co.1.1. Measure Steward: PCPI Foundation

De.3. Brief Description of Measure: Percentage of patients aged 18 years and older with a diagnosis of heart failure (HF) with a current or prior left ventricular ejection fraction (LVEF) < 40% who were prescribed beta-blocker therapy either within a 12-month period when seen in the outpatient setting OR at each hospital discharge

1b.1. Developer Rationale: Beta-blockers are recommended for all patients with stable heart failure and left ventricular systolic dysfunction, unless contraindicated. Treatment should be initiated as soon as a patient is diagnosed with left ventricular systolic dysfunction and does not have low blood pressure, fluid overload, or recent treatment with an intravenous positive inotropic agent. Beta-blockers have been shown to lessen the symptoms of heart failure, improve the clinical status of patients, reduce future clinical deterioration, and decrease the risk of mortality and the combined risk of mortality and hospitalization.

Also, a 2011 analysis of IMPROVE HF data by Fonarow and colleagues revealed that all 4 current ACC/AHA HF outpatient performance measures were associated with decreased risk of 24-month mortality. For the 2 summary measures of HF care processes, there was also a strong positive association between greater conformity to the summary measures and improved risk-adjusted survival. These findings may have significant clinical and public health implications, providing evidence to suggest that current, and some emerging, outpatient process measures may effectively reflect the quality of care provided to patients with HF who are treated in outpatient practice settings.

S.4. Numerator Statement: Patients who were prescribed beta-blocker therapy either within a 12-month period when seen in the outpatient setting OR at each hospital discharge

S.6. Denominator Statement: All patients aged 18 years and older with a diagnosis of heart failure with a current or prior LVEF < 40%

S.8. Denominator Exclusions: Denominator Exceptions:

Documentation of medical reason(s) for not prescribing beta-blocker therapy (e.g., low blood pressure, fluid overload, asthma, patients recently treated with an intravenous positive inotropic agent, allergy, intolerance, other medical reasons).

Documentation of patient reason(s) for not prescribing beta-blocker therapy (e.g., patient declined, other patient reasons).

Documentation of system reason(s) for not prescribing beta-blocker therapy (e.g., other reasons attributable to the healthcare system).

De.1. Measure Type: Process

S.17. Data Source: Electronic Health Records

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

IF Endorsement Maintenance – Original Endorsement Date: Mar 14, 2016 **Most Recent Endorsement Date:** Mar 14, 2016

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Measures #0083 and #0081 (Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) or Angiotensin-Neprilysin Inhibitor (ARNI) Therapy for Left Ventricular Systolic Dysfunction) address related aspects of care for effective treatment for patients with heart failure and should be measured concurrently. Combined treatment with these agents Combined treatment with these agents (i.e., beta-blockers with ACE inhibitor, ARB, or ARNI) produces additive benefits and is required for optimal management of heart failure. It is not recommended that either of these measures be used independently. The pairing of these measures is not intended to suggest the use of any particular scoring methodology (ie, a composite score), nor does it imply either equality of or difference in the relative "weights" of the two measures. A performance score for each measure should be reported individually to provide actionable information upon which to focus quality improvement efforts.

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_0083e_FINAL_08APR19.docx

1a.1 <u>For Maintenance of Endorsement:</u> 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

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0083**e**
Measure Title: Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction
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: 4/9/2019

Instructions

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

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

1a. Evidence to Support the Measure Focus

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

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

Notes

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

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

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one

step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

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

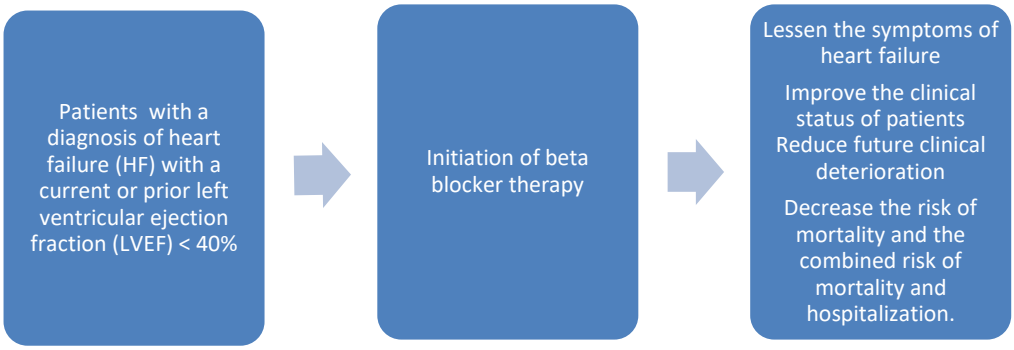
☐ Process: Patients aged 18 years and older with a diagnosis of heart failure (HF) with a current or prior left ventricular ejection fraction (LVEF) < 40% who were prescribed beta-blocker therapy either within a 12-month period when seen in the outpatient setting OR at each hospital discharge

☐ 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.



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.

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

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

☐Clinical Practice Guideline recommendation (with evidence review)

☐US Preventive Services Task Force Recommendation

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

☐Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	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. Available at: https://www.ahajournals.org/doi/full/10.1161/CIR.0b013e31829e8776
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.	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 [heart failure with reduced ejection fraction], unless contraindicated, to reduce morbidity and mortality (Class I, Level of Evidence: A) (ACCF/AHA, 2013). Treatment with a beta blocker should be initiated at very low doses [see excerpt from guideline table below] followed by gradual increments in dose if lower doses have been well tolerated... Clinicians should make every effort to achieve the target doses of the beta blockers shown to be effective in major clinical trials. Even if symptoms do not improve, long-term treatment should be maintained to reduce the risk of major clinical events. Abrupt withdrawal of treatment with a beta blocker can lead to clinical deterioration and should be avoided (ACCF/AHA, 2013). Drugs Commonly Used for Stage C HFrEF (abbreviated to align with focus of measure to include only Beta-blocker therapy) Drug — Initial Daily Dose(s) — Maximum Dose(s) — Mean Doses Achieved in Clinical Trials Beta Blockers Bisoprolol — 1.25 mg once — 10 mg once — 8.6 mg/d Carvedilol — 3.125 mg twice — 50 mg twice — 37 mg/d Carvedilol CR — 10 mg once — 80 mg once — N/A Metoprolol succinate extended release — 12.5 to 25 mg once — 200 mg once — 159 mg/d

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**Level of Evidence A:** Data derived from multiple randomized clinical trials or meta analyses **Level of Evidence B:** Data derived from a single randomized trial, or nonrandomized studies
Provide all other grades and definitions from the evidence grading system	**Level C:** Only consensus opinion of experts, case studies, or standard of care.
Grade assigned to the **recommendation** with definition of the grade	**Class I:** Recommendation that the procedure or treatment is useful/effective
Provide all other grades and definitions from the recommendation grading system	**Class IIa:** Recommendation in favor if treatment or procedure being useful/effective **Class IIb:** Recommendation's usefulness/efficacy less well established **Class III No Benefit:** Procedure/test/treatment is not helpful or has no proven benefit **Class III Harm:** Procedure/test/treatment incurs excess cost without benefit or is harmful to patients
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	17 Randomized Controlled Trials, 3 comparative studies There are many solid randomized controlled trials that show that the benefits of using beta blockers greatly outweigh the harms. They are very effective and relatively safe. The benefits of beta blockers were seen in patients with or without CAD and in patients with or without diabetes mellitus, as well as in women and blacks. The favorable effects of beta blockers were also observed in patients already taking ACE inhibitors.
Estimates of benefit and consistency across studies	Long-term treatment with beta blockers can lessen the symptoms of HF, improve the patient's clinical status, and enhance the patient's overall sense of well-being. In addition, like ACE inhibitors, ARB, or ARNI therapy, beta blockers can reduce the risk of death and the combined risk of death or hospitalization.
What harms were identified?	Initiation of treatment with a beta blocker may produce 4 types of adverse reactions that require attention and management: fluid retention and worsening HF; fatigue; bradycardia or heart block; and hypotension. The occurrence of fluid retention or worsening HF is not generally a reason for the permanent withdrawal of treatment. Such patients generally respond favorably to intensification of conventional therapy, and once treated, they remain excellent candidates for long-term treatment with a beta blocker. The slowing of heart rate and cardiac conduction produced by beta blockers is generally asymptomatic and thus requires no treatment; however, if the bradycardia is accompanied by dizziness or lightheadedness or if second- or third-degree heart block occurs, clinicians should decrease the dose of the beta blocker. Clinicians may minimize the risk of hypotension by administering the beta blocker and ACE inhibitor at different times during the day. Hypotensive symptoms may also resolve after a decrease in the dose of diuretics in patients who are volume depleted. If hypotension is accompanied by other clinical evidence of hypoperfusion, beta-blocker therapy should be decreased or discontinued pending further patient evaluation. The symptom of fatigue is multifactorial and is perhaps the hardest symptom to address with confidence. Although fatigue may be related to beta blockers, other causes of fatigue should be considered, including sleep apnea, overdiuresis, or depression.

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	The articles supporting the beta blocker recommendation were from 1989-2009. However, the overall literature search was through October 2011, with select articles included through April 2013. We ran a search for heart failure and beta blockers for 2014 and 2015. There are several studies related to beta blockers and their use in heart failure. *Nebivolol is not currently recommended for treatment of Heart Failure and is not included in the measure. The 2013 guideline cites one study from 2009 and says "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." Montero et al (2014) does show some benefit, at least in the elderly. We await the next revision of the guideline before considering changes to the measure.* 1) Montero-Perez-Barquero M, Flather M, Roughton M, Coats A, Böhm M, Van Veldhuisen DJ, Babalis D, Solal AC, Manzano L. Influence of systolic blood pressure on clinical outcomes in elderly heart failure patients treated with nebivolol: data from the SENIORS trial. Eur J Heart Fail. 2014 Sep;16(9):1009-15. doi: 10.1002/ejhf.136. Epub 2014 Jul 17. Montero et al (2014) looked at the influence of systolic blood pressure on clinical outcomes in elderly patients with heart failure treated with nebivolol. Patients were divided into three baseline pre-treatment SBP categories (<110, 110-130, and >130 mmHg). They also evaluated the influence of SBP (≤ 130 and > 130 mmHg) on patients with LVEF <40% vs. ≥ 40%. Low baseline SBP was associated with worse clinical outcomes irrespective of treatment group, both in patients with reduced EF and in those with preserved EF. Nebivolol had similar benefits irrespective of baseline SBP: the hazard ratio (HR) for primary outcome of all-cause mortality or cardiovascular hospitalization in the three SBP categories for nebivolol vs. placebo was 0.85 [95% confidence interval (CI) 0.50-1.45], 0.79 (95% CI 0.61-1.01), and 0.88 (95% CI 0.72-1.07), respectively (P for interaction = 0.61). Similar results were obtained for the secondary endpoint of all-cause mortality. There was no significant interaction for the effects of nebivolol by baseline SBP stratified by LVEF. They conclude that elderly HF patients with lower SBP have a worse outcome than those with higher SBP, but nebivolol appears to be safe and well tolerated, with similar benefits on the composite outcome of death or cardiovascular hospital admission irrespective of baseline SBP and LVEF. *Some studies compared cardevilol to metoprolol and bisoprolol- all 3 are currently recommended by the guideline and are included as part of the measure. One study concluded that heart failure patients receiving high-dose carvedilol (≥50 mg daily) showed significantly lower all-cause mortality risk and hospitalization risk, compared with other beta-blockers. This is clearly still an area of interest in the research community. As such, we will wait for the new research to be examined as part of the guideline update process before considering changes to the measure.* 2) Bølling R, Scheller NM, Køber L, Poulsen HE, Gislason GH, Torp-Pedersen C. Comparison of the clinical outcome of different beta-blockers in heart failure patients: a retrospective nationwide cohort study. Eur J Heart Fail. 2014 Jun;16(6):678-84. doi: 10.1002/ejhf.81. Epub 2014 Apr 4. Bolling et al (2014) looked at all Danish patients ≥35 years of age who were hospitalized with a first admission for heart failure and who initiated treatment with a beta-blocker within 60 days of dischargefrom 1995-2011. The main outcome was all-cause mortality and all-cause hospitalization. Cox proportional hazard models were used to compare survival. The study included 58 634 patients of whom 30.121 (51.4%) died and 46.990 (80.1%) were hospitalized during follow-up. The mean follow-up time was 4.1 years. In an unadjusted model carvedilol was associated with a lower mortality [hazard ratio (HR) 0.737, 0.714-0.761] compared with metoprolol (reference) while bisoprolol was not associated with an increased mortality (HR 1.020, 0.973-1.069). In a model adjusted for possible confounders and stratified according to beta-blocker dosages, patients that received high-dose carvedilol (≥50 mg daily) had a lower all-cause mortality risk (HR 0.873, 0.789-0.966) than patients receiving high-dose (≥200 mg daily) metoprolol (reference). High-dose bisoprolol (≥10 mg daily) was associated with a greater risk of death (HR 1.125, 1.004-1.261). High-dose carvedilol was associated with significantly lower all-cause hospitalization risk (HR 0.842, 0.774-0.915) than high-dose metoprolol (reference), while high-dose bisoprolol had insignificantly lower risk than high-dose metoprolol (HR 0.948, 0.850-1.057). They concluded that heart failure patients receiving high-dose carvedilol (≥50 mg daily) showed significantly lower all-cause mortality risk and hospitalization risk, compared with other beta-blockers.

3) Molenaar P, Christ T, Berk E, Engel A, Gillette KT, Galindo-Tovar A, Ravens U, Kaumann AJ. Carvedilol induces greater control of β2- than β 1-adrenoceptor-mediated inotropic and lusitropic effects by PDE3, while PDE4 has no effect in human failing myocardium. Naunyn Schmiedebergs Arch Pharmacol. 2014 Jul;387(7):629-40. doi: 10.1007/s00210-014-0974-4. Epub 2014 Mar 26.

The β-blockers carvedilol and metoprolol provide important therapeutic strategies for heart failure treatment. Therapy with metoprolol facilitates the control by phosphodiesterase PDE3, but not PDE4, of inotropic effects of catecholamines in human failing ventricle. However, it is not known whether carvedilol has the same effect. The authors investigated whether the PDE3-selective inhibitor cilostamide (0.3 μM) or PDE4-selective inhibitor rolipram (1 μM) modified the positive inotropic and lusitropic effects of catecholamines in ventricular myocardium of heart failure patients treated with carvedilol. Right ventricular trabeculae from explanted hearts of nine carvedilol-treated patients with terminal heart failure were paced to contract at 1 Hz. The effects of (-)-noradrenaline, mediated through β1-adrenoceptors (β2-adrenoceptors blocked with ICI118551), and (-)-adrenaline, mediated through β2-adrenoceptors (β1-adrenoceptors blocked with CGP20712A), were assessed in the absence and presence of the PDE inhibitors. The inotropic potency, estimated from -logEC50s, was unchanged for (-)-noradrenaline but decreased 16-fold for (-)-adrenaline in carvedilol-treated compared to non-β-blocker-treated patients, consistent with the previously reported β2-adrenoceptor-selectivity of carvedilol. Cilostamide caused 2- to 3-fold and 10- to 35-fold potentiations of the inotropic and lusitropic effects of (-)-noradrenaline and (-)-adrenaline, respectively, in trabeculae from carvedilol-treated patients. Rolipram did not affect the inotropic and lusitropic potencies of (-)-noradrenaline or (-)-adrenaline. Treatment of heart failure patients with carvedilol induces PDE3 to selectively control the positive inotropic and lusitropic effects mediated through ventricular β2-adrenoceptors compared to β1-adrenoceptors. The β2-adrenoceptor-selectivity of carvedilol may provide protection against β2-adrenoceptor-mediated ventricular overstimulation in PDE3 inhibitor-treated patients. PDE4 does not control β1- and β2-adrenoceptor-mediated inotropic and lusitropic effects in carvedilol-treated patients.

And finally, a meta-analysis analyzed patient data to look at the use of beta blockers in the subgroup of patients with heart failure and atrial fibrillation. They concluded that beta blockers should not be used preferentially over other rate-control medications and not regarded as standard therapy to improve prognosis in patients with concomitant heart failure and atrial fibrillation. Again, we will wait for revised guideline recommendations before considering changes to the measure.

4) Kotecha D, Holmes J, Krum H, Altman DG, Manzano L, Cleland JG, Lip GY, Coats AJ, Andersson B, Kirchhof P, von Lueder TG, Wedel H, Rosano G, Shibata MC, Rigby A, Flather MD; Beta-Blockers in Heart Failure Collaborative Group. Efficacy of β blockers in patients with heart failure plus atrial fibrillation: an individual-patient data meta-analysis.Lancet. 2014 Dec 20;384(9961):2235-43. doi: 10.1016/S0140-6736(14)61373-8. Epub 2014 Sep 2.

Kotecha et al (2014) noted that the efficacy of these drugs in heart failure patients with concomitant atrial fibrillation is ncertain. They meta-analysed individual-patient data to assess the efficacy of β blockers in patients with heart failure and nus rhythm compared with atrial fibrillation.

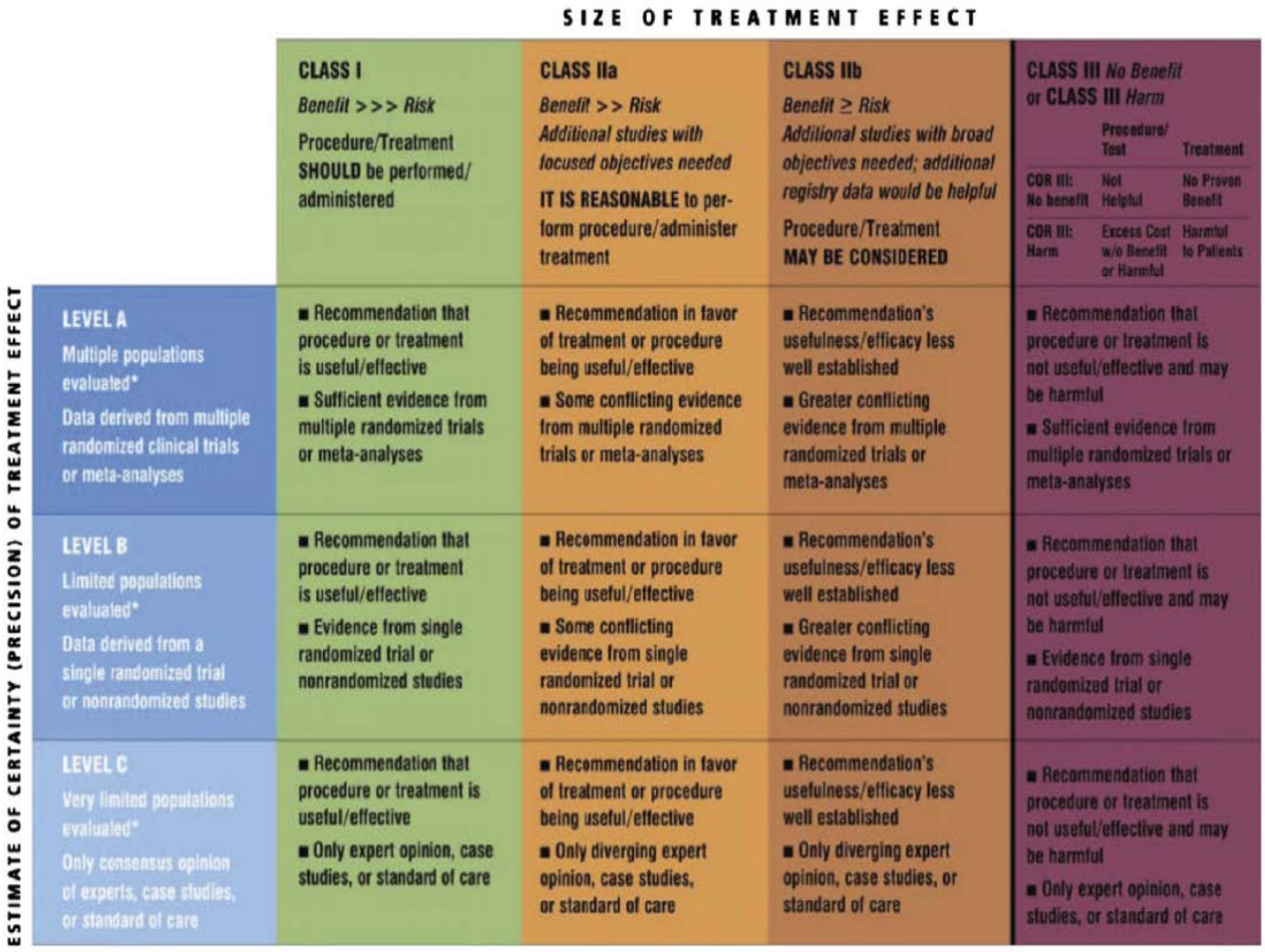
They extracted individual-patient data from ten randomised controlled trials of the comparison of β blockers versus lacebo in heart failure. The presence of sinus rhythm or atrial fibrillation was ascertained from the baseline lectrocardiograph. The primary outcome was all-cause mortality. Analysis was by intention to treat. Outcome data were eta-analysed with an adjusted Cox proportional hazards regression. The study is registered with Clinicaltrials.gov, umber NCT0083244, and PROSPERO, number CRD42014010012.

18,254 patients were assessed, and of these 13,946 (76%) had sinus rhythm and 3066 (17%) had atrial fibrillation at aseline. Crude death rates over a mean follow-up of 1·5 years (SD 1·1) were 16% (2237 of 13,945) in patients with sinus ythm and 21% (633 of 3064) in patients with atrial fibrillation. β-blocker therapy led to a significant reduction in all-ause mortality in patients with sinus rhythm (hazard ratio 0·73, 0·67-0·80; p<0·001), but not in patients with atrial brillation (0·97, 0·83-1·14; p=0·73), with a significant p value for interaction of baseline rhythm (p=0·002). The lack of fficacy for the primary outcome was noted in all subgroups of atrial fibrillation, including age, sex, left ventricular ejection raction, New York Heart Association class, heart rate, and baseline medical therapy.

Based on their findings, they conclude that β blockers should not be used preferentially over other rate-control edications and not regarded as standard therapy to improve prognosis in patients with concomitant heart failure and trial fibrillation.

	While there was a focused update of the guideline that supports this measure in 2014, the specific recommendations that support this measure were not included in the update and remain unchanged. An updated search covering January 1, 2016 through March 31, 2019 was performed. 349 articles were found using the MeSH search terms “Adrenergic beta-Agonists” and “Heart Failure”. However, there were very few studies that are directly applicable to the target population of this measure, and none would change the recommendation to prescribe beta-blocker therapy. **As the measure developer, we would wait until an updated systematic review of the body of evidence is conducted which can confirm or refute the findings of any study published since the guideline was released, considering the full body of evidence available.**

Table 1. Applying Classification of Recommendation and Level of Evidence



ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT	SIZE OF TREATMENT EFFECT: CLASS I	CLASS IIa	CLASS IIb	CLASS III *No Benefit* or CLASS III *Harm*
	Benefit >>> Risk Procedure/Treatment **SHOULD** be performed/ administered	*Benefit >> Risk* *Additional studies with focused objectives needed* **IT IS REASONABLE** to perform procedure/administer treatment	*Benefit ≥ Risk* *Additional studies with broad objectives needed; additional registry data would be helpful* Procedure/Treatment **MAY BE CONSIDERED**	Procedure/Test: COR III: No benefit — Not Helpful; COR III: Harm — Excess Cost w/o Benefit or Harmful. Treatment: No Proven Benefit; Harmful to Patients
LEVEL A Multiple populations evaluated* Data derived from multiple randomized clinical trials or meta-analyses	■ Recommendation that procedure or treatment is useful/effective ■ Sufficient evidence from multiple randomized trials or meta-analyses	■ Recommendation in favor of treatment or procedure being useful/effective ■ Some conflicting evidence from multiple randomized trials or meta-analyses	■ Recommendation's usefulness/efficacy less well established ■ Greater conflicting evidence from multiple randomized trials or meta-analyses	■ Recommendation that procedure or treatment is not useful/effective and may be harmful ■ Sufficient evidence from multiple randomized trials or meta-analyses
LEVEL B Limited populations evaluated* Data derived from a single randomized trial or nonrandomized studies	■ Recommendation that procedure or treatment is useful/effective ■ Evidence from single randomized trial or nonrandomized studies	■ Recommendation in favor of treatment or procedure being useful/effective ■ Some conflicting evidence from single randomized trial or nonrandomized studies	■ Recommendation's usefulness/efficacy less well established ■ Greater conflicting evidence from single randomized trial or nonrandomized studies	■ Recommendation that procedure or treatment is not useful/effective and may be harmful ■ Evidence from single randomized trial or nonrandomized studies
LEVEL C Very limited populations evaluated* Only consensus opinion of experts, case studies, or standard of care	■ Recommendation that procedure or treatment is useful/effective ■ Only expert opinion, case studies, or standard of care	■ Recommendation in favor of treatment or procedure being useful/effective ■ Only diverging expert opinion, case studies, or standard of care	■ Recommendation's usefulness/efficacy less well established ■ Only diverging expert opinion, case studies, or standard of care	■ Recommendation that procedure or treatment is not useful/effective and may be harmful ■ Only expert opinion, case studies, or standard of care

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.

Beta-blockers are recommended for all patients with stable heart failure and left ventricular systolic dysfunction, unless contraindicated. Treatment should be initiated as soon as a patient is diagnosed with left ventricular systolic dysfunction and does not have low blood pressure, fluid overload, or recent treatment with an intravenous positive inotropic agent. Beta-blockers have been shown to lessen the symptoms of heart failure, improve the clinical status of patients, reduce future clinical deterioration, and decrease the risk of mortality and the combined risk of mortality and hospitalization.

Also, a 2011 analysis of IMPROVE HF data by Fonarow and colleagues revealed that all 4 current ACC/AHA HF outpatient performance measures were associated with decreased risk of 24-month mortality. For the 2 summary measures of HF care processes, there was also a strong positive association between greater conformity to the summary measures and improved risk-adjusted survival. These findings may have significant clinical and public health implications, providing evidence to suggest that current, and some emerging, outpatient process measures may effectively reflect the quality of care provided to patients with HF who are treated in outpatient practice settings.

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.*

2016 EHR data from the PQRS program was provided to the PCPI by CMS for the purposes of testing the measure. The data are analyzed for the time period January 2016 through December 2016 and include 52,213 quality events. The mean performance rate is 0.72, the standard deviation is 0.32, the minimum is 0, the maximum is 1, and the interquartile range is 0.50 (1.00 – 0.50). Performance Scores by Decile: (1st,0.10; 2nd,0.50; 3rd,0.62; 4th,0.74; 5th,0.87; 6th,1.00; 7th,1.00; 8th,1.00; 9th,1.00; 10th,1.00)

Historical PQRS data from the PQRS experience report does not differentiate between EHR and Registry average performance rates. Performance scores over time are for 2013: 0.78, 2014: 0.69, 2015: 0.86 It should be noted that PQRS was a voluntary reporting program. Overall participation in the program was suboptimal with 72% of eligible professionals using any method to participate in PQRS, in 2016. The performance scores listed above are not consistently derived from a nationally representative sample.

Quality benchmarks for MIPS 2018 were made publicly available in January 2019. As MIPS is a new program, historical PQRS data was used with MIPS eligibility criteria applied in order to create the benchmark. Providers earn points depending what decile of the benchmark they fall into. The EHR average performance rate reported in the benchmark report is 73.2% and standard deviation of 19.6. Deciles 3 through 10 are also reported and are as follows: Decile, Performance (3rd, 60.00%-65.78%, 4th, 65.79%-70.58%, 5th, 70.59%-74.99%, 6th, 75.00%-79.99%. 7th, 80.00%-85.95%, 8th, 85.96%-90.47%, 9th, 90.48%-95.99%, 10th, =96.00%. While not made explicit in the publicly available documentation, it is thought that deciles 1 and 2 are not

included in the file since providers earn the same amount of points for results in those deciles regardless of performance. No additional data is available at this time.

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.

A study evaluating 15,205 Medicare beneficiaries hospitalized for heart failure with reduced ejection fraction (HFrEF) between 2007 to 2013 examined beta-blocker prescription fill rates after hospital discharge. Researchers found that 38% of hospitalizations were followed by a prescription fill for a beta-blocker within 30 days. While most commonly contraindications (e.g., hypotension, COPD, or syncope) were why the prescription was not filled, that alone does not explain the low a fill rate. While this study focused on prescription fill rates, the prescription has to be written (the focus of this measure) in order for the patient to fill it. (1)

According to Fonarow and colleagues (2010), for aggregate practices at baseline, a ß-blocker was prescribed for 11 868 (86.2%) of 13 772 eligible patients. (2)

1. Loop MS, van Dyke MK, Chen L, Safford MM, Kilgore ML, Brown TM, et al. Low utilization of beta-blockers among Medicare beneficiaries for heart failure with reduced ejection fraction. J Card Fail. 2018. DOI: 10.1016.j.cardfail.2018.10.005

2. Fonarow GC; Albert NM; Curtis AB; Stough WG; Gheorghiade M; Heywood T; McBride M; Inge PJ; Mehra MR; O'Connor CM; Reynolds D; Walsh MN; Yancy CW. Improving Evidence-Based Care for Heart Failure in Outpatient Cardiology Practices: Primary Results of the Registry to Improve the Use of Evidence-Based Heart Failure Therapies in the Outpatient Setting (IMPROVE HF). Circulation 2010; 122: 585-596. Published online before print July 26, 2010, doi: 10.1161/CIRCULATIONAHA.109.934471.

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.*

While this measure is included in several federal reporting programs, those programs have not yet made disparities data available for us to analyze and report.

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

A 2011 study by Bagchi et al of the TRICARE program found that African Americans were less likely than whites to have received beta blockers and angiotensin-converting enzyme inhibitors or angiotensin receptor blockers following a CHF diagnosis ($P<0.0001$). Hispanics were, in some cases, equally likely as whites to receive pharmacological treatments for CHF. In multivariate models, there were no significant racial/ethnic differences in the odds of a potentially avoidable hospitalization (PAH); age greater than 65 was the most significant predictor of a PAH. This study suggests that although there are some racial and ethnic disparities in the receipt of pharmacological therapy for CHF among TRICARE beneficiaries, these differences do not translate into disparities in the likelihood of a PAH. The findings support previous research suggesting that equal access to care may mitigate racial/ethnic health disparities.

Bagchi AD, Stewart K, McLaughlin C, Higgins P, Croghan T. Treatment and outcomes for congestive heart failure by race/ethnicity in TRICARE. Med Care. 2011 May;49(5):489-95. doi: 10.1097/MLR.0b013e318207ef87.

http://www.ncbi.nlm.nih.gov/pubmed/21422958

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 : Congestive Heart Failure

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):*

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.)*

The measure specifications are attached to this submission. Additional measure details may be found at: eCQI Resource Center https://ecqi.healthit.gov/eligible-professional-eligible-clinician-ecqms. Value set details at VSAC: https://vsac.nlm.nih.gov/.

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 an eMeasure **Attachment:** CMS144_v5_6_Artifacts_2019Apr09.zip

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:** 0083e_HF_BetaBlocker_ValueSets_20190409.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.

Supporting guidelines and coding value sets included in the measure are reviewed on an annual basis. This annual review has resulted minor changes to the value sets, to account for updates to the coding terminologies for existing data elements. Measure specifications are annually updated to align with any changes to the standards or tools used to support electronic measurement. Beginning with 2019 implementation, the measure was revised to have two populations: 1.) Patients who were prescribed beta-blocker therapy within a 12-month period when seen in the outpatient setting OR 2.) Patients who were prescribed beta-blocker therapy at each hospital discharge. This change was made to more clearly delineate the denominator requirements to promote accurate implementation. Based on feedback we heard regarding how vendors have implemented the measure, there was an inconsistent approach to applying the measure criteria. Therefore, we decided to split this measure out into two populations, based on the care setting, which can be implemented in both the eCQM and registry versions of this measure. Though the measure is split into two, the measure still requires only one performance rate for reporting.

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 who were prescribed beta-blocker therapy either within a 12-month period when seen in the outpatient setting OR at each hospital discharge

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).

Time Period for Data Collection: At least once during the measurement period when seen in the outpatient setting OR at each hospital discharge

Definition:

Prescribed-Outpatient setting: prescription given to the patient for beta-blocker therapy at one or more visits in the measurement period OR patient already taking beta-blocker therapy as documented in current medication list.

Prescribed-Inpatient setting: prescription given to the patient for beta-blocker therapy at discharge OR beta-blocker therapy to be continued after discharge as documented in the discharge medication list.

Guidance:

Beta-blocker therapy: For patients with prior LVEF < 40%, beta-blocker therapy should include bisoprolol, carvedilol, or sustained release metoprolol succinate.

HQMF eCQM developed and is attached to this submission in fields S.2a and S.2b.

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

All patients aged 18 years and older with a diagnosis of heart failure with a current or prior LVEF < 40%

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).

Time Period for Data Collection: 12 consecutive months

Definition:

LVEF < 40% corresponds to qualitative documentation of moderate dysfunction or severe dysfunction.

Guidance:

A range value should satisfy the logic requirement for ´Ejection Fraction´ as long as the ranged observation value clearly meets the less than 40% threshold noted in the denominator logic. A range that is inclusive of or greater than 40% would not meet the measure requirement.

To satisfy this measure, it must be reported for all heart failure patients at least once during the measurement period if seen in the outpatient setting. If the patient has an eligible inpatient discharge during the measurement period, as defined in the measure logic, it is expected to be reported at each hospital discharge.

The requirement of two or more visits is to establish that the eligible professional or eligible clinician has an existing relationship with the patient

HQMF eCQM developed and is attached to this submission in fields S.2a and S.2b.

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

Denominator Exceptions:

Documentation of medical reason(s) for not prescribing beta-blocker therapy (e.g., low blood pressure, fluid overload, asthma, patients recently treated with an intravenous positive inotropic agent, allergy, intolerance, other medical reasons).

Documentation of patient reason(s) for not prescribing beta-blocker therapy (e.g., patient declined, other patient reasons).

Documentation of system reason(s) for not prescribing beta-blocker therapy (e.g., other reasons attributable to the healthcare system).

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.)*

Time Period for Data Collection: During the encounter within the 12-month period

Exceptions are used to remove a patient from the denominator of a performance measure when the patient does not receive a therapy or service AND that therapy or service would not be appropriate due to patient-specific reasons. The patient would otherwise meet the denominator criteria. Exceptions are not absolute, and are based on clinical judgment, individual patient characteristics, or patient preferences. The PCPI exception methodology uses three categories of reasons for which a patient may be removed from the denominator of an individual measure. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For measure Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD), exceptions may include medical reason(s) (e.g., low blood pressure, fluid overload, asthma, patients recently treated with an intravenous positive inotropic agent, allergy, intolerance, other medical reasons), patient reason(s) (e.g., patient declined, other patient reasons), or system reason(s) (e.g., other reasons attributable to the healthcare system) for not prescribing beta-blocker therapy. Where examples of exceptions are included in the measure language, value sets for these examples are developed

and included in the eCQM. Although this methodology does not require the external reporting of more detailed exception data, the PCPI recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The PCPI also advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

HQMF eCQM developed and is attached to this submission in fields S.2a and S.2b.

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.)*

Consistent with CMS' Measures Management System Blueprint and recent national recommendations put forth by the IOM and NQF to standardize the collection of race and ethnicity data, we encourage the results of this measure to be stratified by race, ethnicity, administrative sex, and payer and have included these variables as recommended data elements to be collected.

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.)*

This measure is comprised of two populations but is intended to result in one reporting rate. The reporting rate is the aggregate of Population 1 and Population 2, resulting in a single performance rate. For the purposes of this measure, the single performance rate can be calculated as follows:

Performance Rate = (Numerator 1 + Numerator 2)/ [(Denominator 1 - Denominator Exceptions 1) + (Denominator 2 - Denominator Exceptions 2)]

Calculation algorithm for Population 1: Patients who were prescribed beta-blocker therapy within a 12-month period when seen in the outpatient setting

1. Find the patients who meet the initial population (i.e., the general group of patients that a set of performance measures is designed to address).

2. From the patients within the initial population criteria, find the patients who qualify for the denominator (i.e., the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial population and denominator are identical.

3. From the patients within the denominator, find the patients who meet the numerator criteria (i.e., the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator.

4. From the patients who did not meet the numerator criteria, determine if the provider has documented that the patient meets any criteria for exception when denominator exceptions have been specified [for this measure: medical reason(s) (e.g., low blood pressure, fluid overload, asthma, patients recently treated with an intravenous positive inotropic agent, allergy, intolerance, other medical reasons), patient reason(s) (e.g., patient declined, other patient reasons), or system reason(s) (e.g., other reasons attributable to the healthcare system) for not prescribing beta-blocker therapy]. If the patient meets any exception criteria, they should be removed from the denominator for performance calculation. -- Although the exception cases are removed from the denominator population for the performance calculation, the exception rate (i.e., percentage 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.

Calculation algorithm for Population 2: Patients who were beta-blocker therapy at each hospital discharge

1. Find the patients who meet the initial population (i.e., the general group of patients that a set of performance measures is designed to address).

2. From the patients within the initial population criteria, find the patients who qualify for the denominator (i.e., the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial population and denominator are identical.

3. From the patients within the denominator, find the patients who meet the numerator criteria (i.e., the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator.

4. From the patients who did not meet the numerator criteria, determine if the provider has documented that the patient meets any criteria for exception when denominator exceptions have been specified [for this measure: medical reason(s) (e.g., low blood pressure, fluid overload, asthma, patients recently treated with an intravenous positive inotropic agent, allergy, intolerance, other medical reasons), patient reason(s) (e.g., patient declined, other patient reasons), or system reason(s) (e.g., other reasons attributable to the healthcare system) for not prescribing beta-blocker therapy]. If the patient meets any exception criteria, they should be removed from the denominator for performance calculation. -- Although the exception cases are removed from the denominator population for the performance calculation, the exception rate (i.e., percentage 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.

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.

Not applicable. The measure is not based on a sample.

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

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

Not applicable. The measure is not based on a survey.

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

If other, please describe in S.18.

Electronic Health 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.

Not applicable

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

No data collection instrument provided

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

Clinician : Group/Practice, Clinician : Individual

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

Home Care, Inpatient/Hospital, Other, Outpatient Services

If other: Domiciliary, Nursing Facility

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.)*

Not applicable. The measure is not a composite.

2. Validity – See attached Measure Testing Submission Form

v2_0083e_nqf_testing_attachment_7.1-636849656960475174.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 Number (*if previously endorsed*)**:** 0083

Measure Title: Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)

Date of Submission: 2/8/2019

Type of Measure:

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

Instructions

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

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

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

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

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

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

Notes

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

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

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

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

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

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

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

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

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

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

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)
The data source is EHR data.

Data 2 (GPRO EHR Web-Interface)
The data source is the Centers for Medicare & Medicaid Service PQRS GPRO EHR Web Interface data base.

Bonnie Patient Test Deck
As a supplement to the EHR reliability testing performed on this measure, a deck of patient test cases has been developed and a summary of the details has been included as part of the feasibility attachment in section 3b.3 of the measure submission form.

Data 3 (GPRO Registry)
See registry submission

Data 4 (EHR – Exceptions Analysis)
The data source is EHR data.

Current Testing

The data source is 2016 EHR data from the PQRS program, provided by the Center for Medicare & Medicaid Services (CMS), and includes data reported from a large number of certified EHR vendors. These vendors include several of the major EHR solutions used by inpatient and outpatient care practices. For example: Allscripts, Epic, MEDITECH, Cerner, GE Healthcare, Nextgen, eClinicalWorks, and other smaller EHR vendors.

In 2016 there were six participation options for submitting measure data to PQRS. Of those, the following can be used to submit EHR data:

- Eligible Providers (EPs) could submit data directly through a qualified EHR product or through a qualified data submission vendor that is Certified EHR Technology.
- Group practices with 2 or more EPs can participate through the group practice reporting option (GPRO) using an EHR direct submission or qualified data submission vendor that is Certified EHR Technology.

To participate, EPs and Group practices submit performance data such as number of eligible instances (denominator), instances of quality service performed (numerator), number of performance exclusions, reporting rates, and performance rates—in a file format specified by CMS. Data is then summarized at the practice level and includes both EPs participating individually as well as group practices participating through GPRO.

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

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)
The data are collected from patients sampled from 2007.
Data 2 (GPRO EHR Web-Interface)
The data are for the time period January 2013 – December 2013, and cover the entire United States.
Data 4 (EHR – Exceptions Analysis)
The data are collected from patients sampled from 2009.

Current Testing

The data are for the time period January 2016 through December 2016 and cover the entire United States. Given the required conversion to ICD-10 in late 2015, the testing was completed on the ICD-10 specified measure.

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*)

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

The data sample came from an academic general internal medicine clinic with several years of experience using a commercial EHR.

Data 2 (GPRO EHR Web-Interface)

The total number of physicians reporting on this measure is 142. Of those, 129 physicians had all the required data elements and met the minimum number of quality reporting events (10) for inclusion in the reliability analysis. For this measure, 90.8 percent of physicians are included in the analysis, and the average number of quality reporting events is 90.1 for a total of 11,628 events. The range of quality reporting events for 129 physicians included is from 553 to 10. The average number of quality reporting events for the remaining 9.2 percent of physicians who aren't included is 4.4.

For this measure, the minimum number of observations for inclusion in signal-to-noise reliability testing was 10 events. Given the structure of the PQRS program, a physician may choose to submit or not submit to PQRS. Since these data contain results on a large number of physicians, limiting the reliability analysis to only those physicians who are participating in the program will eliminate the bias introduced by the inclusion of from physicians who are in the data, but are not submitting to PQRS.

Data 4 (EHR – Exceptions Analysis)

The data sample came from five physician offices using five different EHR systems.

Current Testing

We received data from 3,582 providers reporting on this measure through the EHR reporting option for CMS's PQRS in 2016. This data set reflects a combination of individual provider data and group data and our analysis of the data as a whole is reflected throughout this submission. Of those, 939 providers had all the required data elements and met the minimum number of quality reporting events (10) for a total of 33,877 quality events. For this measure, 26 percent of providers are included in the analysis, and the average number of quality reporting events are 36 for the remaining 33,877 events. The range of quality reporting events for 939 providers included is from 10 to 383. The average number of quality reporting events for the remaining 74 percent of providers that aren't included is 3.

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

The sample consisted of approximately 254 charts for a total of 254 eligible patients. One trained investigator reviewed the 254 charts. The patients were selected using random sampling.

Data 2 (GPRO EHR Web-Interface)

There were 11,628 patients included in this testing and analysis. These were the patients that were associated with physicians who had 10 or more patients eligible for this measure.

Data 4 (EHR – Exceptions Analysis)

The sample consisted of approximately 118 eligible patients.

Current Testing

There were 33,877 patients included in this reliability testing and analysis. These were the patients that were associated with providers who had 10 or more patients eligible for this measure.

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.

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)
The data sample was used for the purposes of reliability and validity testing.

Data 2 (GPRO EHR Web-Interface)
The same data sample from each data source was used for reliability testing and exceptions analysis.

Face Validity (Data 2 & Data 3)
After the measure was fully specified, an expert panel of 12 members was asked to rate their agreement with the following statement:

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

Data 4 (EHR – Exceptions Analysis)
The data sample was used for the exception analysis only.

Current Testing

The same data samples were used for reliability testing and exceptions analysis.

Empirical validity correlation testing was conducted using Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #005).

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.

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)
This was not captured as part of the testing.

Data 2 (GPRO EHR Web-Interface)
This was not captured as part of the testing.

Current Testing

Patient-level socio-demographic (SDS) variables were not captured as part of the testing as that information was not provided in the CMS data used for analysis.

_____

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*)

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

See 2b2.2 for Validity Against the Gold Standard Results

Data 2 & Data 3 (Signal-to-Noise Reliability)

Reliability of the computed measure score was measured as 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 physician performance. Reliability at the level of the specific physician is given by:

Reliability = Variance (physician-to-physician) / [Variance (physician-to-physician) + Variance (physician-specific-error]

Reliability is the ratio of the physician-to-physician variance divided by the sum of the physician-to-physician variance plus the error variance specific to a physician. 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 physician performance.

Reliability testing was performed by using a beta-binomial model. The beta-binomial model assumes the physician performance score is a binomial random variable conditional on the physician's true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.

Reliability is estimated at two different points, at the minimum number of quality reporting events for the measure and at the mean number of quality reporting events per physician.

Current Testing

Reliability of the computed measure score was measured as 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 provider performance and the noise is the total variability in measured performance. Reliability at the level of the specific provider is given by:

Reliability = Variance (provider-to-provider) / [Variance (provider-to-provider) + Variance (provider-specific-error]

Reliability is the ratio of the provider-to-provider variance divided by the sum of the provider-to-provider variance plus the error variance specific to a provider.

Reliability testing was performed by using a beta-binomial model. The beta-binomial model assumes the provider performance score is a binomial random variable conditional on the provider's true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.

Reliability is evaluated by averaging over provider specific reliabilities for all providers that meet the minimum number of quality reporting events for the measure. Each provider must have at least 10 eligible reporting events to be included in this calculation.

A reliability equal to zero implies that all the variability in a measure is attributable to measurement error. A reliability equal to one implies that all the variability is attributable to real differences in provider performance. A reliability of 0.70 – 0.80 is generally considered the acceptable threshold for reliability, 0.80 – 0.90 is considered high reliability, and 0.90 – 1.0 is considered very high. [1]

1. Adams JL, Mehrotra A, McGlynn EA, Estimating Reliability and Misclassification in Physician Profiling, Santa Monica, CA: RAND Corporation, 2010. www.rand.org/pubs/technical_reports/TR863. (Accessed on February 24, 2012.)

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*)

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

See 2b2.3 for Validity Against the Gold Standard Results

Data 2 (GPRO EHR Web-Interface)

For this measure, the reliability at the minimum level of quality reporting events (10) was 0.44. The average number of quality reporting events for physicians included is 90.1. The reliability at the average number of quality reporting events was 0.87.

Current Testing

The reliability above the minimum level of quality reporting events was 0.86. The reliability including providers with less than 10 eligible reporting events is 0.81.

Table 1: Reliability Results

	Previous testing data reliability results	Current testing data reliability results
1+ events	0.87	0.81
10+ events	0.44	0.86

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?*)

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

See 2b2.4 for Validity Against the Gold Standard Results

Data 2 (GPRO EHR Web-Interface)

This measure has moderate reliability when evaluated at the minimum level of quality reporting events and high reliability at the average number of quality events.

Current Testing

This measure has high reliability when evaluated above the minimum level of quality reporting events and high reliability when including providers with less than the minimum level of quality reporting events.

_____

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)*

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

Data abstracted from randomly sampled patient records were used to evaluate parallel forms reliability for the measure. Charts for abstraction were selected for patients aged 18 years and older with heart failure.

Face Validity (Data 2 & Data 3)

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

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

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

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

Current Testing

For this measure, the PCPI has conducted review and updates to the measure specifications, which satisfy the NQF's ICD-10 Conversion requirements. We are providing the information below to support the three requirements:

- NQF ICD-10-CM Requirement 1: Statement of intent related to ICD-10 CM
  Goal was to convert this measure to a new code set, fully consistent with the original intent of the measure.

- NQF ICD-10-CM Requirement 2: Coding Table
  See attachment in S.2b

- NQF ICD-10-CM Requirement 3: Description of the process used to identify ICD-10 codes
  The PCPI uses the General Equivalence Mappings (GEMs) as a first step in the identification of ICD-10 codes. We then review the ICD-10 codes to confirm their inclusion in the measure is consistent with the measure intent, making additions or deletions as needed. We have an RHIA-credentialed professional on our staff who reviews all ICD-10 coding. For measures included in CMS' Quality Payment Program (QPP), the ICD-10 codes have also been reviewed and vetted by the CMS contractor. Comments received from stakeholders related to ICD-10 coding are first reviewed internally. Depending on the nature of the comment received, we also engage clinical experts to advise us as to whether a change to the specifications is warranted.

Validity testing method

Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #008) and Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #005) were chosen as suitable candidates for correlation analysis due to the similarities in patient population and domain. We hypothesize that there exists a positive association of scores between providers who prescribe beta-blocker therapy for patients with a diagnosis of heart failure with a current or prior LVEF < 40% either within a 12 month period when seen in the outpatient setting or at each hospital discharge and those providers who prescribe ACE inhibitor or ARB therapy for patients with a diagnosis of heart failure with a current or prior LVEF < 40% either within a 12 month period when seen in the outpatient setting or at each hospital discharge.

Providers included in the analysis met the minimum number of quality reporting events (10) and were cleaned in the same process as the PQRS dataset.

Datasets were reviewed to identify shared providers based on NPI and TIN identifiers. Correlation analysis was then performed to evaluate the association between performance scores of these shared providers.

We use the following guidance to describe correlation[1]:

Correlation	Interpretation
> 0.40	Strong
0.20 - 0.40	Moderate
< 0.20	Weak

1. Shortell T. An Introduction to Data Analysis & Presentation. Sociology 712. http://www.shortell.org/book/chap18.html. Accessed July 13, 2018.

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

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

Of the 254 patients sampled, automated EHR review detected 219 (86.2%) with an active electronic prescription for a Beta Blocker. Of the remaining 35 patients, 13(37.1%) met one or more of the exclusion criteria. Performance on the Beta Blocker quality measure was 90.9% by using automated EHR review.

Manual review of clinicians' notes in the EHR did not reveal any additional patients who were being treated with a Beta Blocker. However, 6 additional patients who met the exclusion criteria were identified. One patient had met the exclusion criteria through automated review, but upon manual review, was identified to be false.

Performance on the measure was calculated to be 92.8% through comparison of automated and manual EHR review.

Face Validity (Data 2 & Data 3)

Our expert panel included 12 members. Panel members were comprised of experts from the AMA-PCPI Measure Advisory Committee. The list of expert panel members is as follows:

Amy Sanders, MD, MS
David Seidenwurm, MD
Dianne V. Jewell, PT, DPT, PhD, CCS, FAACVPR
Janet Sullivan, MD
John Easa, MD, FIPP
Joseph P. Drozda, Jr., MD, FACC
Mark Metersky, MD
Martha J. Radford, MD, FACC, FAHA
Michael O'Dell, MD, MS, MSHA, FAAFP
Richard Bankowitz, MD, MBA, FACP
Scott T. MacDonald, MD
Shannon Sims, MD, PhD

Current Testing

Data from the PQRS program were used to perform the correlation analysis for this measure. Data comes from the EHR versions of Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #008) and Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #005).

Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #008) was positively correlated with Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #005).

PQRS #005

Coefficient of correlation = 0.65
P-value < 0.001
Number of shared providers based on NPI and TIN identifiers = 872

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?*)

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

The automated quality assessment had a sensitivity of 100.0% for identifying patients with heart failure taking a Beta Blocker. The automated quality assessment captured 12 of 18 patients with valid exclusion criteria (sensitivity, 66.7%), and 1 of 13 patients who met exclusion criteria were judged not to have a true exclusion.

Data 2 and Data 3 (GPRO EHR Web-Interface and GPRO Registry-Face Validity)

The results of the expert panel rating of the validity statement were as follows: N = 12; Mean rating = 4.33 and 100.0% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings
1 – 0 responses (Strongly Disagree)
2 – 0 responses
3 – 0 responses (Neither Agree nor Disagree)
4 – 8 responses
5 – 4 responses (Strongly Agree)

Current Testing

Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #008) has a strong positive correlation with Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #005). The correlation is statistically significant at the 90% significance level and with a coefficient of correlation of 0.65, the correlation is strong. The strong positive correlation with Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD) (PQRS #005) demonstrates the criterion validity of the measure.

_____

2b2. EXCLUSIONS ANALYSIS

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

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

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

This data sample was not used to test exclusions.

Data 2 and Data 3 (GPRO EHR Web-Interface and GPRO Registry)

With the information available from the GPRO Registry and GPRO EHR Web-Interface, we are unable to determine the type of exception reported. However, the exceptions data captured were analyzed to determine frequency and variability across providers.

Data 4 (EHR-Exceptions Analysis)

Exceptions included documentation of medical reason(s), patient reason(s) and system reason(s) for not prescribing beta-blocker therapy. Exceptions were analyzed for frequency and variability across providers.

Current Testing

Exceptions include:

- Documentation of medical reason(s) for not prescribing beta-blocker therapy (eg, low blood pressure, fluid overload, asthma, patients recently treated with an intravenous positive inotropic agent, allergy, intolerance, other medical reasons).
- Documentation of patient reason(s) for not prescribing beta-blocker therapy (eg, patient declined, other patient reasons).
- Documentation of system reason(s) for not prescribing beta-blocker therapy (eg, other reasons attributable to the healthcare system).

Exceptions were analyzed for frequency across providers.

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*)

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

This data sample was not used to test exclusions.

Data 2 (GPRO EHR Web-Interface)

Amongst the 129 physicians with the minimum (10) number of quality reporting events, there were a total of 601 exceptions reported. The average number of exceptions per physician in this sample is 4.7. The overall exception rate is 4.9%.

Data 4 (EHR-Exceptions Analysis)

Reported exceptions were validated upon manual review of the medical record, against an a priori list generated by expert opinion. Measure exceptions were validated 95.32% of the time. Review of the 118 exceptions revealed 98.0% of exceptions were medical reasons for not prescribing beta blocker therapy. Medical reason exceptions consisted of clinical contraindications, drug allergy and drug intolerance.

Current Testing

Amongst the 939 providers with the minimum (10) number of quality reporting events, there were a total of 3,168 exceptions reported. The average number of exceptions per provider in this sample is 3.37. The proportion of exceptions to patients is 0.09.

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*)

Previous 2015 Testing

Exceptions are necessary to account for those situations when it is not medically appropriate to prescribe beta blocker therapy. Exceptions are discretionary and the methodology used for measure exception categories are not uniformly relevant across all measures; for this measure, there is a clear rationale to permit an exception for medical, patient or system reasons. Rather than specifying an exhaustive list of explicit medical, patient or system reasons for exception for each measure, the measure developer relies on clinicians to link the exception with a specific reason for the decision not to prescribe beta blocker therapy required by the measure.

Some have indicated concerns with exception reporting including the potential for physicians to inappropriately exclude patients to enhance their performance statistics. Research has indicated that levels of exception reporting occur infrequently and are generally valid (Doran et al., 2008), (Kmetik et al., 2011). Furthermore, exception reporting has been found to have substantial benefits: "it is precise, it increases acceptance of [pay for performance] programs by physicians, and it ameliorates perverse incentives to refuse care to "difficult" patients." (Doran et al., 2008).

Although this methodology does not require the external reporting of more detailed exception data, the measure developer recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. We also advocate for the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

Without exceptions, the performance rate would not accurately reflect the true performance of that physician. This would result in an increase in performance failures and false negatives. The additional value of increased data collection of capturing an exception greatly outweighs the reporting burden.

References:

Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of pay for performance targets by English Physicians. New Engl J Med. 2008; 359: 274-84.

Kmetik KS, Otoole MF, Bossley H et al. Exceptions to Outpatient Quality Measures for Coronary Artery Disease in Electronic Health Records. Ann Intern Med. 2011;154:227-234.

Current Testing

See previous 2015 testing response above

_____

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.
Previous 2015 Testing

Not applicable

Current Testing

Not applicable

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.
Previous 2015 Testing

Not applicable

Current Testing

Not applicable

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?
Previous 2015 Testing

Not applicable

Current Testing

Not applicable

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?

Previous 2015 Testing

Not applicable

Current Testing

Not applicable

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.**

Previous 2015 Testing

Not applicable

Current Testing

Not applicable

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*)

Previous 2015 Testing

Not applicable

Current Testing

Not applicable

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*)**:**

Previous 2015 Testing

Not applicable

Current Testing

Not applicable

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Previous 2015 Testing

Not applicable

Current Testing

Not applicable

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Previous 2015 Testing

Not applicable

Current Testing

Not applicable

2b3.9. Results of Risk Stratification Analysis:
Previous 2015 Testing

Not applicable

Current Testing

Not applicable

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*)
Previous 2015 Testing

Not applicable

Current Testing

Not applicable

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*)
Previous 2015 Testing

Not applicable

Current Testing

Not applicable

_____

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)*

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

This data sample was not used to test for meaningful differences in performance across providers or practice sites.

Data 2 (GPRO EHR Web-Interface)

Measures of central tendency, variability, and dispersion were calculated.

Current Testing

Measures of central tendency, variability, and dispersion were calculated.

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*)

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

This data sample was not used to test for meaningful differences in performance across providers or practice sites.

Data 2 (GPRO EHR Web-Interface)

Based on the sample of 129 included physicians, the mean performance rate is 0.90, the median performance rate is 0.92 and the mode is 1.00. The standard deviation is 0.09. The range of the performance rate is 0.37, with a minimum rate of 0.63 and a maximum rate of 1. The interquartile range is 0.12 (0.85 - 0.97).

Current Testing

Based on the sample of 939 included providers, the mean performance rate is 0.72, the median performance rate is 0.74 and the mode is 1.00. The standard deviation is 0.21. The range of the performance rate is 0.99, with a minimum rate of 0.01 and a maximum rate of 1.00. The interquartile range is 0.29 (0.89–0.6).

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?*)

Previous 2015 Testing

Data 1 (EHR - Validity Against the Gold Standard)

This data sample was not used to test for meaningful differences in performance across providers or practice sites.

Data 2 (GPRO EHR Web-Interface)

The range of performance from 0.63 to 1.00 suggests there's clinically meaningful variation across physicians' performance.

Current Testing

The range of performance from 0.01 to 1.00 suggests that there exists clinically meaningful variation across providers' performance.

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*)
Previous 2015 Testing

This test was not performed for this measure.

Current Testing

This test was not performed for this measure.

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*)

Previous 2015 Testing

This test was not performed for this measure.

Current Testing

This test was not performed for this measure.

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*)
Previous 2015 Testing

This test was not performed for this measure.

Current Testing

This test was not performed for this measure.

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*)

Previous 2015 Testing

Data are not available to complete this testing.

Current Testing
The PQRS dataset provided to us by CMS did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when submitted to CMS in which case those values would not be counted towards measure performance. There is no indication that this missing data was systematic, thus their omission would lead to unbiased performance results.

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*)
Previous 2015 Testing

Data are not available to complete this testing.

Current Testing

This test was not performed for this measure. There was no missing data.

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*)
Previous 2015 Testing

Data are not available to complete this testing.

Current Testing

The PQRS dataset provided to us by CMS did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when submitted to CMS in which case those values would not be counted towards measure performance. There is no indication that this missing data was systematic, thus their omission would lead to unbiased performance results.

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 health records (EHRs)

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.

We have not identified an areas of concern or made any modifications as a result of testing and operational use of the measure in relation to data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, and other feasibility issues unless otherwise noted.

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 Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial uses of the Measures require a license agreement between the user and the AMA, (on behalf of the PCPI), ACC or AHA.

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.

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	Payment Program Quality Payment Program Merit-Based Incentive Payment System (MIPS) https://qpp.cms.gov/mips/quality-measures Quality Payment Program Merit-Based Incentive Payment System (MIPS) https://qpp.cms.gov/mips/quality-measures

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

1) Merit-based Incentive Payment System (MIPS)-Sponsored by the Centers for Medicare and Medicaid Services (CMS)

Prior to 2016, this measure was used for Eligible Providers (EPs) in the Physician Quality Reporting System (PQRS). As of 2017, PQRS has been replaced by the Merit-based Incentive Payment System (MIPS). MIPS is a national performance-based payment program that uses performance scores across several categories to determine payment rates for EPs. MIPS takes a comprehensive approach to payment by basing consideration of quality on a set of evidence-based measures that were primarily developed by clinicians, thus encouraging improvement in clinical practice and supporting advances in technology that allow for easy exchange of information.

According to the CY 2019 Quality Payment Program final rule, CMS intends to "make all measures under MIPS quality performance category available for public reporting on Physician Compare in the transition year of the Quality Payment Program, as technically feasible." These measures include those reported via all available submission methods for MIPS-eligible clinicians and groups. Because this measure has been in use for at least one year and meets the minimum sample size requirement for reliability, this measure meets criteria for public reporting. 2018 data will be available for public reporting on Physician Compare in late 2019. The Registry version of this measure is currently included in the downloadable database on the Physician Compare website and is not yet available on individual or group profiles.

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?*)

We support the expanded use of this measure in government or other programs, including those intended for accountability or public reporting. The ACC, AHA and PCPI do not have any policies that would restrict access to the performance measure specifications or results or that would impede implementation of the measure for any application. We would welcome its implementation in emerging applications such as accountable care organizations (ACO), Medicare Advantage insurance plans or health plans selling on the insurance marketplace.

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.*)

As described above, it is our understanding that CMS is also planning to move towards publicly reporting physician data via Physician Compare. This measure is currently included in the downloadable database on the Physician Compare website and is not yet available on individual or group profiles. Also, although the measure is currently in use, we support expanded use of this measure in government or other programs, including those intended for accountability or public reporting.

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

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

The PCPI measure development and maintenance process is a rigorous, evidence-based process that has been refined and standardized since the PCPI's inception in 2000. Throughout its tenure, the PCPI has conducted its measure development and maintenance process with strict adherence to several key principles, including the following which underscore the role those being measured have played in the development and maintenance process and in providing feedback based on measure implementation:

Collaborative Approach to Measure Development

PCPI measures are developed and maintained through cross-specialty, multi-disciplinary technical expert panels. Representatives of relevant clinical specialties are invited to participate in our expert panels to advise

us throughout the measure development process and as questions arise during measure implementation. Additionally, other health care providers and stakeholders participate in our panels as equal contributors to the measure development process. The PCPI also strives to include on its panels individuals representing the perspectives of patients, consumers, private health plans, and employers. Liaisons from key measure development organizations, including The Joint Commission and NCQA, at times participate in the PCPI's measure development process to ensure measure harmonization. Measure methodologists and coding and informatics experts are also considered important members of the expert panel. This broad-based approach to measure development maximizes the input from those being measured and other stakeholders to develop evidence-based, feasible and clinically meaningful measures.

Public Comment Period

Input from a wide range of stakeholders is integral to the measure development process. To invite other perspectives and expertise beyond the expert panels and particularly from those providers and facilities that will implement these measures, the PCPI submits the measures for public comment. All measures are released for a 30-day public and PCPI member comment period. All comments are reviewed by the technical expert panel to determine whether measure modifications are needed based on comments received.

Feedback Mechanisms

The PCPI has a dedicated mechanism set up to receive measure-related comments and questions from implementers. As comments and questions are received, they are shared with appropriate staff for follow up. If comments or questions require expert input, these are shared with the PCPI's technical expert panels to determine if measure modifications may be warranted. Additionally, for PCPI measures included in federal reporting programs, there is a system that has been set up to elicit timely feedback and responses from PCPI staff in consultation with technical expert panel members, as appropriate.

Feasibility Assessments

The PCPI solicits feedback on measure feasibility in the following domains: data availability, data accuracy, data standards, and workflow to guide future modifications to the measure. During this process, we may receive recommendations to improve the experience of those implementing and reporting on this measure and we follow up on any questions or concerns received by those completing the feasibility assessment. Doing so addresses any issues with interpretation and serves as an important step in the measure development process.

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.

See description in Section 4a2.1.1 above

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.

As described in Section 4a2.1.1, the PCPI invites feedback through various mechanisms. We obtain input from our topic-specific technical expert panels during the measure development and during the annual maintenance process. Additionally, the PCPI obtains feedback via an online public comment and an email-based process set up to receive measure inquiries from implementers.

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

We have received no feedback from those being measured that resulted in any changes to this measure.

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

Based on feedback received via the ONC Project Tracking System, we received a request to clarify the difference between the two populations in this measure as well as explanation of calculation of the single

performance rate, Based on this feedback, a clarifying guidance statement was added to the specifications for the purposes of clarity and consistency. This statement was added in 2018 and was effective for 2019 reporting.

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.

See summary in 4a.2.2.3.

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 PCPI creates measures with an ultimate goal of improving the quality of care, measurement is a mechanism to drive improvement but does not equate with improvement. Measurement can help identify opportunities for improvement with actual improvement requiring making changes to health care processes and structure. In order to promote improvement, quality measurement systems need to provide feedback to front-line clinical staff in as close to real time as possible and at the point of care whenever possible. (1)

1. Conway PH, Mostashari F, Clancy C. The future of quality measurement for improvement and accountability. JAMA. 2013 Jun 5;309(21):2215-6.

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.

The intent of this measure is to improve care of patients diagnosed with heart failure. CMS data report an improvement in performance rates in the last 6 years. However, performance rates represent but one facet of the quality improvement process.

While the PCPI creates measures with an ultimate goal of improving the quality of care, measurement is a mechanism to drive improvement but does not equate with improvement. Measurement can help identify opportunities for improvement with actual improvement requiring making changes to health care processes and/or structure. In order to promote improvement, quality measurement systems need to provide feedback to front-line clinical staff in as close to real time as possible and at the point of care whenever possible. (1)

1. Conway PH, Mostashari F, Clancy C. The future of quality measurement for improvement and accountability. JAMA. 2013 Jun 5;309(21):2215-6.

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

As the prescription of beta blocker therapy for patients with HF who have who have LVEF <40% is part of the pharmacotherapy piece of guideline directed medical therapy (along with prescription of ACE, ARB, or ARNI

therapy), it could be anticipated that rates of prescribing these therapies as well as providing other guideline directed medical therapies would show improvement as well

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%)

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

0083 : Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)

0117 : Beta Blockade at Discharge

0127 : Preoperative Beta Blockade

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

Measure 0083 addresses a therapy which is also covered in part by the following NQF-endorsed measures: NQF 0071: Persistence of Beta-Blocker Treatment After a Heart Attack and NQF 0070 and 0070e: Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%). As a result, the denominator specifications for the measures differ where needed based on the differing patient populations. Additionally, NQF 0071 is intended for use at the health plan level. NQF 0117 is an inpatient/hospital level measure and includes only patients who have undergone isolated CABG surgery. NQF 0127 is also an inpatient/hospital level measure that focuses on administration of beta-blockers prior to isolated CABG surgery.

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.

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): PCPI Foundation

Co.2 Point of Contact: Samantha, Tierney, samantha.tierney@thepcpi.org, 312-224-6071-

Co.3 Measure Developer if different from Measure Steward: PCPI Foundation

Co.4 Point of Contact: Kerri, Fei, kerri.fei@thepcpi.org, 312-224-6070-

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.

PCPI measures are developed and maintained under the aegis of topic-specific technical expert panels (TEPs). The PCPI TEPs are comprised of clinicians and other healthcare professionals representing medical specialty societies and other stakeholders. The TEPs provide clinical expertise as well as advise on methodologic questions and review the measures annually to ensure accuracy and adherence to the most current evidence.

Cardiovascular Technical Expert Panel

Sarah J. Goodlin MD, FACC, FAAHPM (Co-Chair)

Ileana L. Piña MD, MPH (Co-Chair)

Donald E. Casey MD, MPH, MBA

Ted Ganiats MD

Kathleen L. Grady PhD, RN, FAAN

Richard Hellman MD, FACP, FACE

Tony Hermann

Denise M. Kolanczyk PharmD, BCPS-AQ Cardiology

Frederick A. Masoudi MD, MSPH

Joseph V. Messer MD, MACC

David S. Nilasena MD, MSPH, MS

Stephen D. Persell MD, MPH

Paul D. Rockswold MD, MPH, FAAFP

Nancy K. Sweitzer MD, PhD

Carmen M. Terzic MD, PhD

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2003

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

Ad.4 What is your frequency for review/update of this measure? Supporting guidelines and specifications for this measure are reviewed on an annual basis.

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

Ad.6 Copyright statement: Copyright 2019 American College of Cardiology, American Heart Association and American Medical Association. All Rights Reserved.

Ad.7 Disclaimers: The Measure is not a clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications.

The Measure, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain.

Commercial uses of the Measure require a license agreement between the user and the American Medical Association (AMA), American College of Cardiology (ACC) or the American Heart Association (AHA). Neither the AMA, nor ACC, nor AHA, nor the former AMA-convened Physician Consortium for Performance Improvement(R) (AMA-PCPI), nor the PCPI (R) Foundation (PCPI[R]), nor their members shall be responsible for any use of the Measure.

AMA and PCPI encourage use of the Measure by other health care professionals, where appropriate.

THE MEASURE AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

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. The AMA, ACC, AHA, the PCPI and its members and former members of the AMA-PCPI disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT[R]) or other coding contained in the specifications.

CPT(R) contained in the Measure specifications is copyright 2004-2018 American Medical Association. LOINC(R) is copyright 2004-2018 Regenstrief Institute, Inc. This material contains SNOMED Clinical Terms(R) (SNOMED CT[R]) copyright 2004-2018 International Health Terminology Standards Development Organisation. ICD-10 is copyright 2018 World Health Organization. All Rights Reserved.

Due to technical limitations, registered trademarks are indicated by (R) or [R].

Ad.8 Additional Information/Comments: Zip file containing feasibility results for 3b.3 will be sent via email as it cannot be uploaded.