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 #: 0067
Measure Title: Chronic Stable Coronary Artery Disease: Antiplatelet Therapy
Measure Steward: American College of Cardiology
Brief Description of Measure: Percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who were prescribed aspirin or clopidogrel.
Developer Rationale: Improvement in the number of patients with CAD who are prescribed antiplatelet therapy.

Numerator Statement: Patients who were prescribed* aspirin or clopidogrel within a 12 month period.

*Prescribed may include prescription given to the patient for aspirin or clopidogrel at one or more visits in the measurement period OR patient already taking aspirin or clopidogrel as documented in current medication list.
Denominator Statement: All patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period.
Denominator Exclusions: Documentation of medical reason(s) for not prescribing aspirin or clopidogrel (e.g., allergy, intolerance, receiving other thienopyridine therapy, receiving warfarin therapy, bleeding coagulation disorders, other medical reasons)

Documentation of patient reason(s) for not prescribing aspirin or clopidogrel (e.g., patient declined, other patient reasons)

Documentation of system reason(s) for not prescribing aspirin or clopidogrel (e.g., lack of drug availability, other reasons attributable to the health care system)

Measure Type: Process
Data Source: Electronic Clinical Data : Registry
Level of Analysis: Clinician : Individual
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: 8/10/09 **Most Recent Endorsement Date:** 1/18/12

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

0067 Chronic stable coronary artery disease: Antiplatelet therapy
Public and Member Comment
Comments included:
- Concern with broad exclusions.
- Data collection will be difficult for health plans.
- Overlaps with measure 0068 which is in wide use in the private sector.
- Composite measure 0076 is superior to this individual measure.

• Wording should be changed to anti-platelet therapy rather than aspirin or clopidogrel.

Developer Response: The level of analysis for this measure is individual clinician and groups, not health plans. The measure is limited to the only anti-platelet agents (i.e., aspirin and clopidogrel) recommended by the ACC/AHA clinical practice guidelines for patients with chronic stable angina which served as the primary evidence base to support measure development.

Steering Committee: The Committee reviewed the comments and developer responses and again considered the issue of competing measures (0068 & 0076). Ultimately the Committee identified the measures as "overlapping" rather than competing. The Committee identified the narrow population (CAD rather than IVD) as a weakness of this measure.

Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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.

The developer provides the following evidence for this process measure:

- This is a clinician level measure that calculates the percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12-month period who were prescribed aspirin or clopidogrel in a clinician office/clinic, nursing facility or home health.
- The developer provides decision logic from secondary prevention to outcome for the use of antiplatelet therapy in decreasing morbidity, mortality, and hospitalization with patients in with chronic stable CAD.
- The developer provides 4 separate guidelines with 10 guideline statements for use of aspirin and clopidogrel in patients with CAD and provides the details of the Quantity/Quality/Consistency for the 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable IHD, which was Graded Class 1a & 1b and Levels of Evidence of A & B utilizing 2010 ACCF/AHA Task Force on Practice Guidelines. Methodology Manual and Policies From the ACCF/AHA.
- One meta-analysis for STEMI patients was published after the publication of the 2012 discussed guideline comparing intravenous P2Y12 inhibitors with clopidogrel on major ischemic and bleeding events, though the developer states does not conflict with the 10 guideline recommendation statements.

Questions for the Committee:

- *Should patients post-PCI with drug-eluding stents have BOTH aspirin and clopidogrel?*
- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- The developer provided 2013 and 2014 performance data from the Pinnacle registry:

	mean	Std dev	Min	Max	Interquartile score	#providers	#patients	Male	White	Black	Other
2013	86.2%	10.5%	0%	100%	10.3%	2407	1,023,530	62.6%	92.9%	5.1%	2.0%
2014	86.3%	9.49%	0%	100%	10.2%	2248	959,792	62.4%	92.0%	5.8%	2.2%

- Four additional studies cited demonstrating no significant improvement in the number of patients receiving antiplatelets; data demonstrates limited progress on the performance of this measure from 77% in 2006 to 84.9% using PINNACLE Registry data from 2008 to 2009 and 86.2% and 86.3% in 2013 and 2014.
- Disparities performance data from PINNACLE is detailed for total patients and providers based on sex, age. The developer states that race and insurance part of the PINNACLE data collection tool. The sample populations list diabetes, hypertension, Atrial Fibrillation/Flutter, Heart Failure, Peripheral Arterial Disease, Stroke/TIA and Myocardial Infarction co-morbidities.

Questions for the Committe:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

Committee pre-evaluation comments

Criterion 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - This is a process measure that has uses 4 separate guidelines and 10 guideline statements to support the use of aspirin/clopidogrel in patients with CAD. For the guideline for stable IHD the recommendations are graded as 1a & 1b. The levels of of evidence are A (multiple RCTs) and B (single RCT).
   - Extensive/Adequete

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- Process - looks indirectly at ASA/Clopidogrel use via prescription/medicaton listing -- not actual patient use; however given that ASA is available without a prescription, it is as close a proxy as you can get. With Clopidogrel one could look at prescriptions filled but this is still a proxy for patient adherence.
- Yes

1b. ***Committee's Comments on Performance Gap:***

- Information was provided from the Pinnacle registry. This demonstrates a performance gap. There was 86.2% prescribing in 2013 and 86.3% in 2014. In 2009 the rate was 84.9%. This begs the question regarding whether this measure has topped out.
- I am concerned about that fact that the data is all from the Pinnacle Registry where the population was 92.0 - 92.9% White. If there are disparities related to race, they would not necessarily be picked up in the analysis provided. That being said, within the data reported from Pinnacle, there is about a 6% difference in performance rates between men and women patients. The difference between White and Blacks was smaller at around 3%
- Disparities are now well demonstrated
- There is little gap present as the administration of ASA is widely adopted

1c. ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

- Yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure's data source is the PINNACLE Registry with the data dictionary and collection tool provided. ICD9, ICD10, and CPT codes are provided for numerator and denominator. The concepts of "prescribed" and "documented" are provided. Data used to calculate the measure is abstracted electronically from the EHR.
- Based on the title, the measure targets adult patients with "chronic" and "stable" CAD, though neither of these terms are described based on the duration of the diagnosis nor the time from and acute event until measurement. The ICD 9 codes for the denominator include codes for AMI (410) (including those less than 8 weeks old) and 411 other acute and subacute forms of ischemic heart disease (411) (includes acute coronary insufficiency and subendocardia ischemia).
- In 1.2, the developer states that practices with less than 10 patient encounters are excluded from the measure, though this is not included in the denominator exclusions.
- The developer describes both denominator exclusions (#3) and exceptions (#5) in the measure logic calculation in S.18, though only denominator exclusions are provided S.10. Generally, exceptions are noted in eMeasures, which this measure is not specified.
- The developer provides both narrative descriptions for patient, medical and system reason exclusions/exceptions in the submission, though the data collection tool has undefined patient, medical and system exclusions/exceptions.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor(s) review (if not an eMeasure, delete this section): This is not an eMeasure

Prior to 2014, this measure was previously specified as an EHR-reportable measure for PQRS, though it does not meet NQF's current definition of an eMeasure. The developers did not submit an eMeasure with this submission.

2a2. Reliability Testing Testing attachment

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.

- Reliability testing was conducted on the performance score, using PINNACLE data source at the individual clinician level during calendar year 2013 for 2407 providers and 1,023,530 patients; & 2014 for 2248 providers and 959,792 patients.
- A signal-to-noise analysis using the beta-binomial model was conducted. This type of analysis, which is appropriate for the measure, quantifies the amount of variation in performance that is due to differences between providers (as opposed to differences that are due to random measurement error). 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. The method results in a reliability statistic for each clinician.

- At the minimum number of patient visits required (>10) the average reliability was 0.994 for 2013 & 0.986 for 2014. For providers with the median number of patient encounters, the reliability was even higher, with 0.998 for both years. A reliability of 0.70 is generally considered a minimum threshold for acceptability, and 0.80 is considered very good reliability. Very high quartiles reliability statistics are also provided for both 2013 and 2014.

Questions for the Committee:

- *Specific questions on the method and results of reliability testing.*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The clinical practice guidelines supporting this measure recommend the use of aspirin or clopidogrel in patients with CAD.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- Content validity was assessed by an expert work group, public comment, concurrent formal peer review process, approval by the ACC Board of Trustee & Science Advisory & Coordination Committees, and PCPI membership.
- The developers state that construct validity (the degree to which the measure is assessing what it claims to be measuring) was difficult to assess as independent auditing has not occurred, though as the data is abstracted from the EHR by direct transfer, errors would occur by mapping or incorrect auditor abstracting.
- Face validity was systematically assessed via survey by 2 committees (one ACC & one AHA) of 42 members with 83.3% agreeing the measure scores as specified provide an accurate reflection of quality and can be used to distinguish good and poor quality. Using a Likert Scale from 1-5, the mean importance rate was 4.26.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Are further definitions and timing required with the concepts of "chronic" and "stable" for CAD?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer is encouraged to provide clarification if patient, medical and system exclusions/exceptions are directly calculated from a "yes/no" question, or by calculating performance from numerous data elements. Example: Patients excluded from the measure include patients with documented medical (e.g., allergy, intolerance, receiving other thienopyridine therapy, receiving warfarin therapy, bleeding coagulation disorders, other medical reasons), patient (e.g., patient declined, other patient reasons), or system reason (e.g., lack of drug availability, other reasons attributable to the health care system) for not on aspirin or clopidogrel or patients were on warfarin.

- In the PINNACLE data collection tool, the definition and use of patient, medical and system exclusion/exception reasons appear to be left to the provider and/or the EHR vendor mapping the data elements to the clinical registry.
- In PINNACLE missing values are interpreted as "no" for most variables, therefore missing documentation of antiplatelet medication indicates a failure to meet the measure. Providers with less than 10 eligible patient encounters during the study period are also excluded due to small sample sizes. The developers state they did not conduct an empirical analysis of frequency or distribution of missing data.
- There is a wide variation between providers in their rate of exclusions/exceptions.
    - In the 2013 testing sample 3.1% of providers had no exceptions, among those who did, the exclusion rate ranges from 0.2% to 35.3% (mean = 6.0%). Of those patients removed from the measure, 0.72% were removed due to a medical reason, 9.47% were removed for a patient reason, 1.48% were removed for a system reason, 0.30% were removed due to multiple reasons, and 88.03% were removed due to patients who were on warfarin or another thienopyridine that is not included in this measure (e.g. prasugrel or ticagrelor) and the physician felt that the addition of aspirin or a clopidogrel provided an increased risk of bleeding with minimal benefits in the prevention of recurrent cardiovascular events.
    - In the 2014 testing sample 2.5% of the providers had no exceptions, among those who did, the exclusion rate ranges from 0.3% to 27.4% (mean = 5.8%). Of those patients removed from the measure, 0.80% were removed due to a medical reason, 9.52% were removed for a patient reason, 1.88% were removed for a system reason, 0.31% were removed due to multiple reasons, and 87.50% were removed due to concomitant medications that either provide anti-platelet or systemic anti-coagulation.

Questions for the Committee:

- *For the 2013 & 2014 performance and testing data, were all the patients abstracted from EHRs only? If the measure calculates performance electronically abstracted data only, is the population limited only to patients whose providers have an EHR?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This process measure is not risk adjusted.

2b5. Meaningful difference:

- The overall mean performance on this measure is 86.2% (SD = 10.5%) for 2013 and 86.3% (SD = 9.49%) for 2014.
- The developer notes a small amount of variability across providers for statistically 'identical' patients and suggests that a patient presenting to 1 provider, as opposed to another, would on average, be 11% (2013) and 9% (2014) more likely to be treated with an anti-platelet agent.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- The developer notes that in the PINNACLE registry, most missing values are interpreted as "NO".
- Also see missing values.

Committee pre-evaluation comments

Criterion 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1.: ***Committee's Comments on Reliability-Specifications:***

- If this is supposed to be about stable IHD, the codes for AMI (410) and acute and subacute IHD should not be included. There are measures that look at ASA and thienopyridine use in these populations. Exceptions - I do not understand why there would be systems exclusions for this measure.
- The patients that are receiving dual anti platelet therapy for another reason (DAPT) might be included in the numerator - these are technically not chronic stable CAD patients
    - other anti-platelets are not included - prasugrel, ticagrelor
    - I'm not sure that the definition of chronic CAD does specifically exclude patients who had an acute event within the period observed with this measure
    - Patients receiving ASA for other indication might me incorrectly counted as receiving ASA for CAD

2a2.: ***Committee's Comments on Reliability-Testing:***

- Signal-to-noise analysis was performed. The value was very good when the patient visits were >10 (0.994 and 0.986).
- The test sample is adequate. Also the developers provided information about antiplatelet therapy monitoring by other groups and showed similar results
- Yes

2b1.: ***Committee's Comments on Validity-Specifications***:

- The specifications are somewhat broad in that they do not specify a dose or frequency of the antiplatelet therapy. The drugs recommended are consistent with the guidelines.
- Content and face validity were tested as they pertain to the use of antiplatelet therapy in stable IHD.
- It would be reasonable to drop AMI and other acute/subacute IHD from the diagnosis codes in the denominator or put a time stipulation on them to be more reflective of stable outpatients.
- No

2b2.: ***Committee's Comments on Validity-Testing:***

- The developers primarily refer to the Pinnacle Registry testing information; however in a separate testing results document, they provide comparisons concerning antiplatelet medication prescription in other monitoring programs. The results are similar.
- I do believe that the score would be a meaure of quality of stable IHD care. It might be good to put time stipulations on the use of AMI in the denominator although if the patient is being seen in the outpatient setting, it must be at least a few days post-event in most cases.
- Yes

2b3-7.: ***Committee's Comments on Threats to Validity:***

- 2b3 Exclusions - The biggest exclusion was for warfarin use and this is totally appropriate. As mentioned previously, I cannot think of a system reason for not prescribing either ASA or clopidogrel to these patients.
- It does appear that for the performance and testing data, only patients with EHRs were used. Pinnacle can be used without an EHR but requires personnel for data entry.
- 2b4 Risk adjustment - not done; ASA is cheap and easily attainable but it bothers me that the Pinnacle Registry patients were 92% White.
- There was not a lot of variability of performance across the various types of insurance. In fact, those without insurance had higher rates of having prescriptions."
- The patients that are receiving dual anti platelet therapy for another reason (DAPT) might be included in the numerator - these are technically not chronic stable CAD patients
    - other anti-platelets are not included - prasugrel, ticagrelor
    - I'm not sure that the definition of chronic CAD does specifically exclude patients who had an acute event within the period observed with this measure
    - Patients receiving ASA for other indication might me incorrectly counted as receiving ASA for CAD

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- The data source is electronic abstraction from the clinical record, MDS, OASIS to the PINNACLE Registry from readily available data occurring during patient care.
- As one of the earliest NQF endorsed measures utilized, the developers state the data collection strategy is implemented, though no information for abstraction time and costs are provided.

Questions for the Committee:

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

Committee pre-evaluation comments
Criterion 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- The required data is readily available in patient records. Finding the reason for an exception might be more challenging.
- The prescription and diagnosis data elements are readily available. Rationale for not prescribing could be in text fields or written notes making them less accessible. Linking of EHR to Pinnacle extraction would potentially address these issues.
- For the measure to be used outside the Pinnacle registry there could be some data extraction issues.
- The measure is feasible.

Criterion 4: Usability and Use

4. Usability and 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.

- The measure is currently reported in Physician Quality Reporting System (PQRS). Beginning in 2015, a payment adjustment to providers who do not satisfactorily report data on quality measures will be implemented. The measure will also be reported to CMS by the PINNACLE Registry as part of PQRS in 2015 as a Qualified Clinical Data Registry.
- The measure is currently used in the ACC PINNACLE Registry for quality improvement.
- The developers state they continue to seek additional opportunities for measure use in other public reporting and quality improvement programs, and do not have policies that would restrict use.
- In 2014 the Measure Applications Partnership (MAP) Clinician Workgroup did not support the measure for the Physician Compare and Value-Based Payment Modifier Programs because the measure does not adequately address any current needs of the program. Included in a MAP family of measures. Preferring other outcome measures that address coronary artery disease. In 2015, the Clinician Workgroup did not support the measure for the Medicare Shared Savings Program as Optimal Vascular Care measure (E0076) contains this measure as a component of the composite. Both measures would be redundant.

***Questions for the Committee**:*

- *Is the measure publicly reported?*
- *For maintenance measures – is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments

Criterion 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- Not currently publically reported.
- Measure is reported in the PQRS system and within Pinnacle itself
- Multiple other quality monitoring programs use antiplatelet therapy prescription in their plans.
- Unintended consequences are problematic
- It appears that the plateau has been reached and no further improvements are possible
- Public reporting might penalize providers although additional increase in the population receiving ASA is unlikely to happen

Criterion 5: Related and Competing Measures

- 0465 : Perioperative Anti-platelet Therapy for Patients undergoing Carotid Endarterectomy
- 0964 : Therapy with aspirin, P2Y12 inhibitor, and statin at discharge following PCI in eligible patients
- The developer states that the measure specifications are not completely harmonized because 0465 and 0964 address a different patient demographic and focus on inpatient prescribed ASA or clopidogrel.

0067 Chronic Stable Coronary Artery Disease: Antiplatelet Therapy.
(Clinician level measure, used for patients with CAD in OP, NH & HH settings)

Competing Measures
NQF# - Title & Description

Related Measures
NQF# - Title & Description– how it differs – If not in CV, state the other project

Pre-meeting public and member comments

Comment by: Dr. Kathy Gans-Brangs, PhD
Organization: SPI

Comment#5119: "We urge adding Brilinta (ticagrelor) to the specification for NQF# 0067. BRILINTA is an FDA approved P2Y12 platelet inhibitor indicated to reduce the rate of thrombotic cardiovascular (CV) events in patients with acute coronary syndrome (ACS) when given with maintenance doses of aspirin less than 100 mg. In patients treated with percutaneous coronary intervention (PCI), it also reduces the rate of stent thrombosis. The EHR specifications, version 2.0 measure Value Set ID 000200 through 000208 (final 2 pages) include the following drug code descriptions: Thienopyridine therapy-excluding clopidogrel and specifically lists prasugrel, Effient, Ticlopidine and

Ticlid. The measure list does not include Brilinta (ticagrelor).

Supporting Information: The safety and efficacy of BRILINTA was evaluated in PLATO, a multicenter, randomized, double-blind study comparing ticagrelor to clopidogrel in 18,624 patients with ACS.1,2 At 12 months, the rate of CV death/MI/stroke was 9.8% for ticagrelor versus 11.7% for clopidogrel resulting in a relative risk reduction of 16% (p<0.001). The difference between treatments was driven by CV death and MI with no difference in stroke. The relative risk reduction of CV death was 21% and MI was 16% for ticagrelor versus clopidogrel (p=0.0013 and p=0.0045, respectively).1,2 In PLATO, 11,289 (60.6%) patients either had a previous stent implanted (n=1404) or underwent stent implantation during the study (n=9885).7 There was a lower risk of stent thrombosis with ticagrelor (1.3% for adjudicated "definite") than with clopidogrel (1.9%) (hazard ratio [HR], 0.67; 95% CI, 0.50-0.91; p=0.009).1,2,3 The results were similar for drug-eluting stents and bare metal stents.3 The reduction in definite stent thrombosis with ticagrelor was numerically greater for late [> 30 days: HR 0.48, (CI 0.24 – 0.96)], and subacute [24 h – 30 days: HR 0.60, (CI 0.39 – 0.93)] vs. acute stent thrombosis [< 24 h: HR 0.94 (CI 0.43 – 2.05)].

1) BRILINTA Prescribing Information

2) Wallentin L, Becker RC, Budaj A, et al for the PLATO Investigators. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. N Engl J Med. 2009;361:1045-1057

3) Steg PG, Harrington RA, Emanuelsson H, et al. Stent thrombosis with ticagrelor versus clopidogrel in patients with acute coronary syndromes: an analysis from the prospective, randomized PLATO trial. Circulation. 2013;128:1055-1065

Please refer to the BRILINTA Prescribing Information for Boxed Warnings related to increased risk of bleeding and reduced effectiveness with maintenance doses of ASA greater than 100 mg per day (https//www.brilinta.com)."

Comment by: Ashish R. Trivedi, Pharm.D.
Organization: SPI-Lilly
Comment#5114: "While Lilly is supportive of this measure, we suggest the addition of lipid lowering therapy to this measure as supported by treatment guidelines for patients with coronary artery disease (CAD) [Stone et al, 2013]. Also, we would like to point out that comprehensive and routine lipoprotein lipid assessment is still an integral part of managing risk in patients with ASCVD (including CAD) [Jacobson et al, 2015]. In addition, clinical trial data indicates significant residual cardiovascular risk in ASCVD patients treated with statins, even in the setting of optimal LDL-C reduction (eg, <70 mg/dL and <100 mg/dL), thus highlighting the need to consider alternative CV risk reduction algorithms beyond the focus on LDL-C levels and/or the use of statins [Cannon et al 2004, LaRosa et al 2005, Pedersen et al 2005].
References

- Stone NJ, Robinson JG, Lichtenstein AH, et al. ACC/AHA Prevention Guideline: 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129:25 suppl 2 S1-S45, doi:10.1161/01.cir.0000437738.63853.7a
- Jacobson TA, Ito MK, Maki KC, et al. National Lipid Association recommendations for patient-centered management of dyslipidemia: part 1 – Full Report. J Clin Lipidol.2015; 9(2), 129–169. DOI: http://dx.doi.org/10.1016/j.jacl.2015.02.003
- Cannon CP, Braunwald E, McCabe CH, et al. Intensive versus moderate lipid lowering with statin after acute coronary syndromes. N Engl J Med. 2004; 350:1495–1504.
- LaRosa JC, Grundy SM, Waters DD, et al. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. N Engl J Med. 2005; 352:1425–1435.
- Pedersen TR, Faergeman O, Kastelein JJ, et al. High-dose atorvastatin vs usual-dose simvastatin for secondary prevention after myocardial infarction: the IDEAL study: a randomized controlled trial. JAMA. 2005; 294:2437–2445."

Comment by: Dr. Kathy Gans-Brangs, PhD
Organization: SPI
Comment#5118: "REQUEST FOR HARMONIZATION OF SIMILAR MEASURES: We believe that reviews undertaken by NQF in 2013-2014 and 2015 present an opportunity to ensure measure specification drug lists are current – that they exclude obsolete drug products based on inactive NDC codes and include all relevant FDA approved products. We urge the committee to review a side-by-side table of the specification for NQF measure # 0067 with measures 0964, 2452 and 2379 and any other relevant measures to ensure that the P2Y12 platelet inhibitor agents included are consistent (see Measure Comment Report for Cardiovascular Project 2013, Comment Period from May 27, 2014 to June 25, 2014)."

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0067
Measure Title: Chronic Stable Coronary Artery Disease: Antiplatelet Therapy
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: 6/23/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Antiplatelet therapy for patients with chronic stable coronary artery disease

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

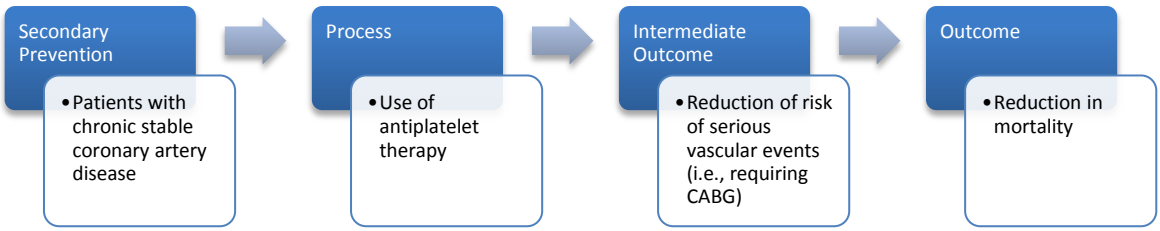
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



Use of antiplatelet therapy can decrease morbidity, mortality, and hospitalizations for patients with chronic stable coronary artery disease.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?
☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***
☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION
1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Smith SC Jr., Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, Gibbons RJ, Grundy SM, Hiratzka LF, Jones DW, Lloyd-Jones DM, Minissian M, Mosca L, Peterson ED, Sacco RL, Spertus J, Stein JH, Taubert KA. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation. 2011: published online before print November 3, 2011, 10.1161/CIR.0b013e318235eb4d.

http://content.onlinejacc.org/article.aspx?articleid=1147807

Amsterdam EA, Wenger NK, Brindis RG, Casey DE Jr, Ganiats TG, Holmes DR Jr, Jaffe AS, Jneid H, Kelly RF, Kontos MC, Levine GN, Liebson PR, Mukherjee D, Peterson ED, Sabatine MS, Smalling RW, Zieman SJ. 2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes: a report of the American College of Cardiology/ American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014;64:e139–228.

http://content.onlinejacc.org/article.aspx?articleid=1910086

O'Gara PT, Kushner FG, Ascheim DD, Casey DE Jr, Chung MK, de Lemos JA, Ettinger SM, Fang JC, Fesmire FM, Franklin BA, Granger CB, Krumholz HM, Linderbaum JA, Morrow DA, Newby LK, Ornato JP, Ou N, Radford MJ, Tamis-Holland JE, Tommaso CL, Tracy CM, Woo YJ, Zhao DX. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;61:e78–140, doi:10.1016/j.jacc.2012.11.019.

http://content.onlinejacc.org/article.aspx?articleid=1486115

Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, Douglas PS, Foody JM, Gerber TC, Hinderliter AL, King

SB III, Kligfield PD, Krumholz HM, Kwong RYK, Lim MJ, Linderbaum JA, Mack MJ, Munger MA, Prager RL, Sabik JF, Shaw LJ, Sikkema JD, Smith CR Jr, Smith SC Jr, Spertus JA, Williams SV. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2012;60:e44–164.

http://content.onlinejacc.org/article.aspx?articleid=1391404

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update (p. e2434)

1. Aspirin 75–162 mg daily is recommended in all patients with coronary artery disease unless contraindicated. **Class I: Level of Evidence: A**
2. Clopidogrel 75 mg daily is recommended as an alternative for patients who are intolerant of or allergic to aspirin. **Class I: Level of Evidence: B**

2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes (p. e161, 171, 175, 187)

p. e161:

3. Non–enteric-coated, chewable aspirin (162 mg to 325 mg) should be given to all patients with NSTE-ACS without contraindications as soon as possible after presentation, and a maintenance dose of aspirin (81 mg/d to 325 mg/d) should be continued indefinitely. **Class I: Level of Evidence: A**

4. In patients with NSTE-ACS who are unable to take aspirin because of hypersensitivity or major gastrointestinal intolerance, a loading dose of clopidogrel followed by a daily maintenance dose should be administered. **Class I: Level of Evidence: A**

p. e171:

5. After PCI, aspirin should be continued indefinitely at a dose of 81 mg to 325 mg daily. **Class I: Level of Evidence: B**

p. e175:

6. Aspirin should be continued indefinitely. The maintenance dose should be 81 mg daily in patients treated with ticagrelor and 81 mg to 325 mg daily in all other patients. **Class I: Level of Evidence: A**

p. e187:

7. Patients with prior CABG and NSTE-ACS should receive antiplatelet and anticoagulant therapy according to GDMT and should be strongly considered for early invasive strategy because of their increased risk. **Class I: Level of Evidence: B**

2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction (p. e91)

8. After PCI, aspirin should be continued indefinitely. **Class I: Level of Evidence: A**

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease (p. e95)

9. Treatment with aspirin 75 to 162 mg daily should be continued indefinitely in the absence of contraindications in patients with SIHD. **Class I: Level of Evidence: A**

10. Treatment with clopidogrel is reasonable when aspirin is contraindicated in patients with SIHD. **Class I: Level of Evidence: B**

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Guideline Statement # (see 1a.4.2 above)	Class of Recommendation/Level of Evidence (for definitions see 1a.4.4 below)
1	Class Ia
2	Class Ib
3	Class Ia
4	Class Ia
5	Class Ib
6	Class Ia
7	Class Ib
8	Class Ia
9	Class Ia
10	Class Ib

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Class of Recommendation (COR) is an estimate of the size of the treatment effect considering risks versus benefits in addition to evidence and/or agreement that a given treatment or procedure is or is not useful/effective or in some situations may cause harm.

Class I: Procedure/Treatment should be performed/administered

Class 1a
- Recommendation that procedure or treatment is useful/effective
- Sufficient evidence from multiple randomized trials or meta-analyses

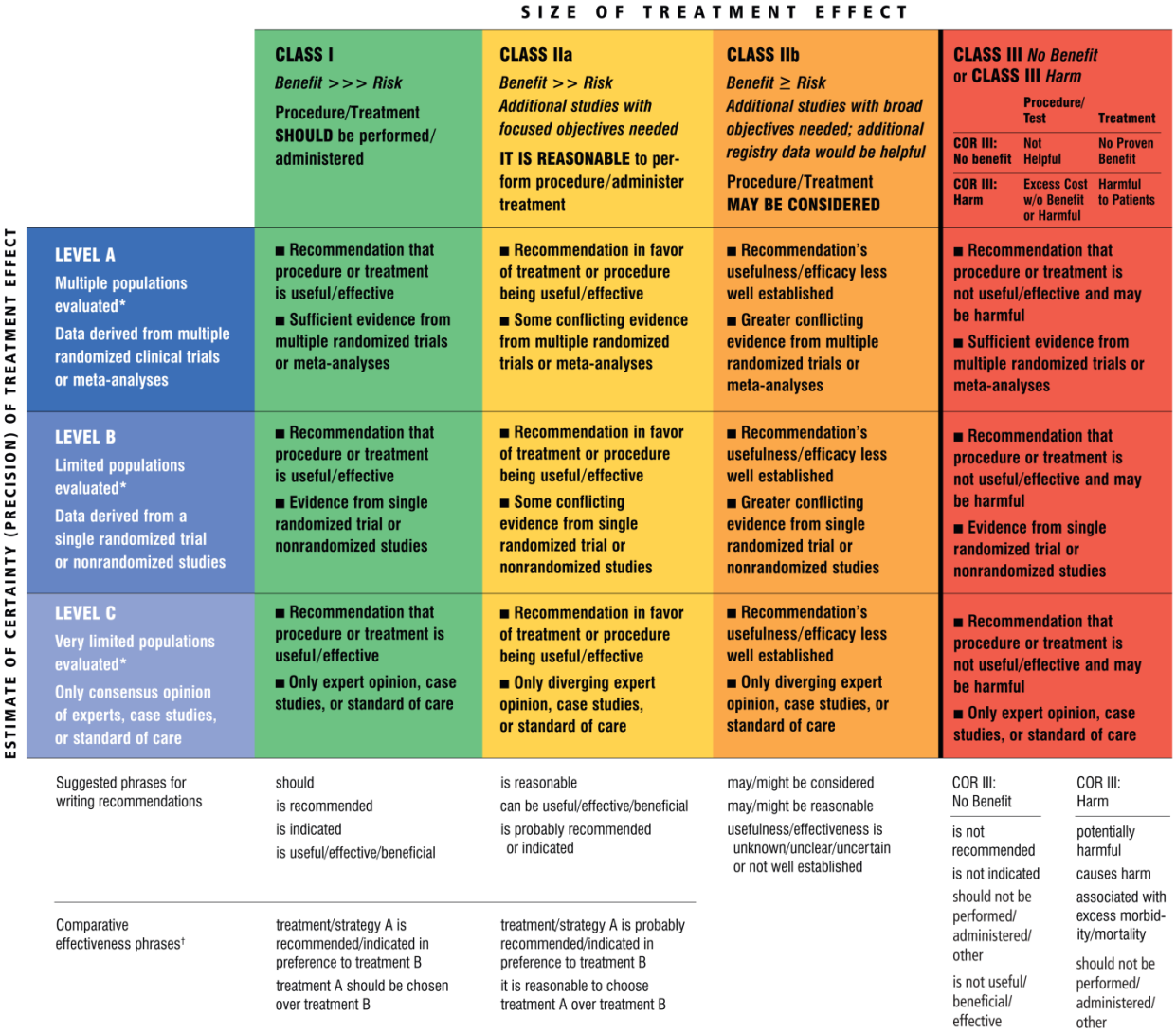
Class 1b
- Recommendation that procedure or treatment is useful/effective
- Evidence from single randomized trial or nonrandomized studies

Class 1c
- Recommendation that procedure or treatment is useful/effective
- Only expert opinion, case studies, or standard of care

Class IIa: It is reasonable to perform procedure/administer treatment
Class IIb: Procedure/Treatment may be considered
Class III: No benefit (Not helpful or No proven benefit)
Class III: Harm (Excess cost w/o benefit or Harmful to patients)

Specific COR definitions are included in Table 1 below.

Table 1. Applying Classification of Recommendation and Level of Evidence



Note: A recommendation with Level of Evidence B or C does not imply that the recommendation is weak. Many important clinical questions addressed in the guidelines do not lend themselves to clinical trials. Although randomized trials are unavailable, there may be a very clear clinical consensus that a particular test or therapy is useful or effective. *Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use. †For comparative effectiveness recommendations (Class I and IIa; Level of Evidence A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

ACCF/AHA Task Force on Practice Guidelines. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines. American College of Cardiology Foundation and American Heart Association, Inc. Cardiosource.com. 2010. Available at: http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf and http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease

This guideline covers multiple management issues for the adult patient with stable known or suspected ischemic heart disease (SIHD) including the guideline-directed medical therapy (GDMT) such as antiplatelet therapy.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

An overall grade for the quality of evidence was not assigned. Rather, the quality of a study (or set of studies) supporting a recommendation was graded on an estimate of the certainty or precision of the treatment effect (see 1a.4.3).

Recommendations used to support this measure have a:

- Level of Evidence of A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation

OR

- Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies. References used to determine level of evidence must be provided and cited with the recommendation.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.
Level of Evidence (LOE) is an estimate of the certainty or precision of the treatment effect.

Level of Evidence A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence C: Consensus opinion of experts, case studies, or standard of care.

Specific LOE definitions are included in Table 1 in 1a.4.4.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: An extensive evidence review was conducted through December 2008 and includes selected other references through December 2011.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease

The body of evidence supporting the recommendations on antiplatelet therapy with patients with a prior MI includes randomized controlled trials and meta-analyses. The number of which is not provided in the guideline.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease

All of the recommendations for this process are rated as Level of Evidence A or B, meaning that the data was derived from one or more RCTs or meta-analyses. Additional information on the overall quality of evidence across the RCTs is not provided.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease

p. e95

Among 2,920 patients with SIHD, a comprehensive meta-analysis of source data revealed an association of aspirin use with a 37% reduction in the risk of serious vascular events, including a 46% decrease in the risk for UA and a 53% decrease in the risk of requiring coronary angioplasty. (Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. BMJ. 2002;324:71– 86.)

Clopidogrel 75 mg has been compared with aspirin 325 mg in patients with previous MI, stroke, or symptomatic PAD in the prospective, randomized CAPRIE (Clopidogrel versus Aspirin in Patients at Risk of Ischaemic Events) study. (A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). CAPRIE Steering Committee. Lancet. 1996;348:1329 –39.) Although clopidogrel demonstrated superiority over aspirin in the secondary prevention of MI and death in this group of patients, the magnitude of difference was small. Because no additional trials comparing aspirin and clopidogrel in patients with SIHD have been conducted, clopidogrel remains an acceptable alternative agent to aspirin.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease

The guideline does not discuss potential harms of aspirin or clopidogrel therapy alone but the potential risks for major bleeding when both therapies are prescribed is discussed.

p. e95

In a meta-analysis of 5 RCTs comparing clopidogrel plus aspirin to aspirin alone in patients with IHD, the incidence of all-cause mortality, MI, and stroke was found to be reduced in the clopidogrel-plus- aspirin group, whereas the risk of major bleeding increased significantly. The incidence of all-cause mortality was 6.3% in the aspirin plus clopidogrel group versus 6.7% in the aspirin group (odds ratio [OR] 0.94; 95% CI 0.89, 0.99; $p = 0.026$). The incidence of myocardial infarction was 2.7% and 3.3% (OR 0.82; 95% CI 0.75, 0.89; $p < 0.0001$), and stroke was 1.2% and 1.4% (OR 0.82; 95% CI 0.73, 0.93; $p = 0.002$). Similarly, the incidence of major bleeding was 1.6% and 1.3% (OR 1.26; 95% CI 1.11, 1.41; $p < 0.0001$), and fatal bleeding was 0.28% and 0.27% (OR 1.04; 95% CI 0.76, 1.43; $p = 0.79$). (Helton TJ, Bavry AA, Kumbhani DJ, et al. Incremental effect of clopidogrel on important outcomes in patients with cardiovascular disease: a meta-analysis of randomized trials. Am J Cardiovasc Drugs. 2007;7:289 –97.)

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

One meta-analysis was published after the publication of the 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease.

Note: Text below for description and results is verbatim from the article abstract.

Tang XF, Fan JU, Meng J, Jin C, Yan JQ, Yang YJ. Impact of new oral or intravenous P2Y12 inhibitors and clopidogrel on major ischemic and bleeding events in patients with coronary artery disease: a meta-analysis of randomized trials. Atherosclerosis. 2014;233:568-78.

Description and Results: Twelve randomized, placebo-controlled studies and two subgroup analyses of included studies on ST-segment elevation myocardial infarction (STEMI) were included. The database consisted of 82,784 patients, with 43,875 (53%) on new oral P2Y12 inhibitors and 38909 (47%) on intravenous P2Y12 inhibitors compared with clopidogrel. The primary efficacy endpoint was major adverse cardiac events (MACEs). The primary safety endpoint was thrombolysis in myocardial infarction (TIMI) major bleeding. New oral P2Y12 inhibitors significantly decreased MACEs (odds ratio: 0.85, $p<0.0001$ for the whole cohort; OR: 0.77, $p=0.04$ for STEMI) and all-cause death (OR: 0.88, $p=0.04$ for the whole cohort; OR: 0.77, $p=0.01$ for STEMI). Among new intravenous P2Y12 inhibitors, only cangrelor significantly decreased the risk of MACEs. An increase in TIMI major bleeding was observed only by prasugrel among the new P2Y12 inhibitors.

Conclusion: New oral P2Y12 inhibitors reduce ischemic events, but there is no obvious increase in major bleeding in patients with CAD, and the risk/benefit ratio is particularly favorable for STEMI patients. Moreover, only cangrelor is beneficial for ischemic events in patients on new intravenous P2Y12 inhibitors.

Impact on conclusions of systematic review: This additional meta-analysis does not impact the current guideline recommendations on which this measure is based.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form CAD_Antiplatelet_0067_Evidence_Form_06_22_15.pdf
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., the benefits or improvements in quality envisioned by use of this measure)*
Improvement in the number of patients with CAD who are prescribed antiplatelet therapy.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
2013 performance data from the Pinnacle registry.

Overall mean performance on this measure is 86.2%, with a standard deviation of 10.5%. The minimum score equals 0.00%, while the maximum score equals 100.00%. The interquartile score is equal to 10.3%.

2,407 providers were measured, and the patient study sample equals 1,023,530. 62.4% of the sample is male. 92.0% of the sample is white, 5.8% is black, and 2.2 % identified as "other." The sample reached across all US regions, with 12.7% of providers in the Northeast, 29.0% of providers in the Midwest, 39.7 % of providers in the South, and 18.6% of providers in the West.
Mean

Decile 1 62.6%
Decile 2 77.7%
Decile 3 82.6%
Decile 4 85.5%
Decile 5 87.6%
Decile 6 89.5%
Decile 7 91.2%
Decile 8 93.0%
Decile 9 94.9%
Decile 10 97.6%

2014 performance data from the Pinnacle registry.

Overall mean performance on this measure is 86.3%, with a standard deviation of 9.49%. The minimum score equals 0.00%, while the maximum score equals 100.00%. The interquartile score is equal to 10.2%.

2,248 providers were measured, and the patient study sample equals 959,792. 62.6% of the sample is male. 92.9% of the sample is white, 5.1% is black, and 2.0% identified as "other." The sample reached across all US regions, with 11.4% of providers in the Northeast, 28.5% of providers in the Midwest, 40.3 % of providers in the South, and 19.8% of providers in the West.

Mean
Decile 1 65.4%
Decile 2 77.8%
Decile 3 82.4%
Decile 4 85.1%
Decile 5 87.3%
Decile 6 89.2%
Decile 7 91.0%
Decile 8 92.8%
Decile 9 94.7%
Decile 10 97.4%

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.

Additional data in support to the importance of CAD antiplatelet use. One study found 77% of patients were prescribed antiplatelet therapy, 51.8% were prescribed BBs, 49.5% were prescribed ACEIs, and 76.9% were prescribed statins [1]. Follow-up assessments 4 years later found no significant change in rates [2]. Most recently, Chan et al. [3], using 2008 to 2009 data from the PINNACLE registry, found higher rates of secondary prevention medication prescription among CAD patients. In their analysis, 84.9% of patients were receiving anti-platelets, 86.4% received BB, 72.4% received ACEI/ARB, and 84.3% received lipid lowering agents [3].

[1] Bhatt DL, Steg PG, Ohman EM, et al. International prevalence, recognition, and treatment of cardiovascular risk factors in outpatients with atherothrombosis. JAMA 2006;295:180–9.

[2] Bhatt DL, Eagle KA, Ohman EM, et al. Comparative determinants of 4-year cardiovascular event rates in stable outpatients at risk of or with atherothrombosis. JAMA 2010;304:1350–7.

[3] Chan PS, Oetgen WJ, Buchanan D, et al. Cardiac performance measure compliance in outpatients: the American College of Cardiology and National Cardiovascular Data registry's PINNACLE (Practice Innovation and Clinical Excellence) program. J Am Coll Cardiol 2010;56: 8–14.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

We examined variation in provider performance on this measure based on sex, age, race and a number of other patient factors to identify variations. The findings are represented for 2013 and 2014 respectively.

To see the data tables in formatted fashion, see testing form page 15.

2013

2013 stratified descriptive statistics of performance rate from Pinnacle Registry.

Male
# of providers: 2403
# of patients: 637419
Minimum: 0.00%
Lower Quartile: 85.7%
Mean: 88.6%
Upper Quartile: 94.7%
Maximum: 100%
Quartile Range: 9.02%
Std Dev: 9.97%

Female
# of providers: 2403
# of patients: 384062
Minimum: 0.00%
Lower Quartile: 77.1%
Mean: 82.5%
Upper Quartile: 90.8%

Maximum: 100%
Quartile Range: 13.8%
Std Dev: 12.4%

Age: <60
# of providers: 2404
# of patients: 197507
Minimum: 0.00%
Lower Quartile: 73.9%
Mean: 81.6%
Upper Quartile: 92.7%
Maximum: 100%
Quartile Range: 18.7%
Std Dev: 15.1%

Age: 60 -< 70
# of providers: 2406
# of patients: 285197
Minimum: 0.00%
Lower Quartile: 83.9%
Mean: 87.5%
Upper Quartile: 94.5%
Maximum: 100%
Quartile Range: 10.6%
Std Dev: 11.1%

Age: 70 -< 80
# of providers: 2405
# of patients: 320245
Minimum: 0.00%
Lower Quartile: 84.6%
Mean: 87.9%
Upper Quartile: 94.4%
Maximum: 100%
Quartile Range: 9.80%
Std Dev: 10.6%

Age: >= 80
# of providers: 2400
# of patients: 220581
Minimum: 0.00%
Lower Quartile: 83.0%
Mean: 87.1%
Upper Quartile: 94.4%
Maximum: 100%
Quartile Range: 11.4%
Std Dev: 11.3%

Insurance: None
# of providers: 134
# of patients: 1381
Minimum: 0.00%

Lower Quartile: 92.3%
Mean: 86.1%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 7.69%
Std Dev: 28.6%

Insurance: Private
# of providers: 683
# of patients: 100517
Minimum: 0.00%
Lower Quartile: 83.9%
Mean: 87.5%
Upper Quartile: 96.0%
Maximum: 100%
Quartile Range: 12.1%
Std Dev: 13.8%

Insurance: Medicaid
# of providers: 500
# of patients: 40963
Minimum: 0.00%
Lower Quartile: 84.7%
Mean: 89.5%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 15.3%
Std Dev: 13.1%

Insurance: Medicare
# of providers: 27
# of patients: 182
Minimum: 50.0%
Lower Quartile: 86.7%
Mean: 92.3%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 13.3%
Std Dev: 12.5%

Insurance: Other
# of providers: 152
# of patients: 820
Minimum: 0.00%
Lower Quartile: 68.3%
Mean: 81.4%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 31.7%
Std Dev: 29.2%

Race: White

# of providers: 1552
# of patients: 502521
Minimum: 0.00%
Lower Quartile: 83.7%
Mean: 86.8%
Upper Quartile: 93.7%
Maximum: 100%
Quartile Range: 10.1%
Std Dev: 10.9%

Race: Black
# of providers: 1425
# of patients: 31466
Minimum: 0.00%
Lower Quartile: 77.8%
Mean: 83.9%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 22.2%
Std Dev: 19.9%

Race: Other
# of providers: 1190
# of patients: 12014
Minimum: 0.00%
Lower Quartile: 75.0%
Mean: 83.8%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 25.0%
Std Dev: 25.6%

2014 stratified descriptive statistics of performance rate from Pinnacle Registry.

2014

Male
# of providers: 2243
# of patients: 599619
Minimum: 0.00%
Lower Quartile: 85.7%
Mean: 88.8%
Upper Quartile: 94.5%
Maximum: 100%
Quartile Range: 8.80%
Std Dev: 8.84%

Female
# of providers: 2242
# of patients: 357647
Minimum: 0.00%
Lower Quartile: 76.6%
Mean: 82.2%
Upper Quartile: 90.6%
Maximum: 100%
Quartile Range: 14.0%
Std Dev: 11.7%

Age: <60
# of providers: 2246
# of patients: 179194
Minimum: 0.00%
Lower Quartile: 74.7%
Mean: 82.0%
Upper Quartile: 92.1%
Maximum: 100%
Quartile Range: 17.3%
Std Dev: 14.2%

Age: 60 -< 70
# of providers: 2248
# of patients: 265376
Minimum: 0.00%
Lower Quartile: 83.7%
Mean: 87.5%
Upper Quartile: 94.1%
Maximum: 100%
Quartile Range: 10.5%
Std Dev: 10.0%

Age: 70 -< 80
# of providers: 2248
# of patients: 306202
Minimum: 20.0%
Lower Quartile: 83.9%
Mean: 87.9%
Upper Quartile: 94.3%
Maximum: 100%
Quartile Range: 10.4%
Std Dev: 9.50%

Age: >= 80
# of providers: 2242
# of patients: 209020
Minimum: 0.00%
Lower Quartile: 83.0%
Mean: 87.1%
Upper Quartile: 93.9%
Maximum: 100%

Quartile Range: 10.9%
Std Dev: 10.5%

Insurance: None
# of providers: 104
# of patients: 528
Minimum: 0.00%
Lower Quartile: 100%
Mean: 90.5%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 0.00%
Std Dev: 24.1%

Insurance: Private
# of providers: 606
# of patients: 99826
Minimum: 0.00%
Lower Quartile: 84.3%
Mean: 88.6%
Upper Quartile: 96.5%
Maximum: 100%
Quartile Range: 12.1%
Std Dev: 12.2%

Insurance: Medicaid
# of providers: 397
# of patients: 21467
Minimum: 0.00%
Lower Quartile: 85.0%
Mean: 89.8%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 15.0%
Std Dev: 13.4%

Insurance: Medicare
# of providers: 27
# of patients: 177
Minimum: 0.00%
Lower Quartile: 80.0%
Mean: 87.9%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 20.0%
Std Dev: 21.5%

Insurance: Other
# of providers: 103
# of patients: 284
Minimum: 0.00%
Lower Quartile: 77.8%

Mean: 83.0%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 22.2%
Std Dev: 31.8%

Race: White
# of providers: 1288
# of patients: 428912
Minimum: 0.00%
Lower Quartile: 83.7%
Mean: 87.1%
Upper Quartile: 93.3%
Maximum: 100%
Quartile Range: 9.58%
Std Dev: 9.79%

Race: Black
# of providers: 1212
# of patients: 23387
Minimum: 0.00%
Lower Quartile: 77.2%
Mean: 84.7%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 22.8%
Std Dev: 18.4%

Race: Other
# of providers: 1090
# of patients: 9331
Minimum: 0.00%
Lower Quartile: 76.9%
Mean: 84.6%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 23.1%
Std Dev: 25.4%

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
We are not aware of any publications/evidence outlining disparities in this area, but would refer the reader to information provided in 1b.4.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

•Coronary heart disease (CHD) alone caused about 1 of every 6 deaths in the United States in 2010. In 2010, 379, 559 Americans died of CHD. [1]

•Each year, an estimated 620,000 Americans have a new coronary attack (defined as first hospitalized myocardial infarction or coronary heart disease death) and about 295, 000 have a recurrent attack. [2]

•It is estimated that an additional 150,000 silent first myocardial infarctions occur each year. Approximately every 34 seconds, 1 American has a coronary event, and approximately every 1 minute 23 seconds, an American will die of one.[2]

•Between 2013 and 2030, medical costs of CHD (real 2010$) are projected to increase by approximately 100% Indirect costs for all CVD (real 2010$) are projected to increase 52% (from $202.5 billion to $308.2 billion) between 2013 and 2030. Of these indirect costs, CHD is projected to account for about 43% and has the largest indirect costs (AHA computation, based on methodology described by Heidenreich et al.[3]

•On the basis of the NHLBI-sponsored FHS CHD makes up more than half of all cardiovascular events in men and women <75 years of age. The incidence of CHD in women lags behind men by 10 years for total CHD and by 20 years for more serious. [4]

•In 2006, $11.7 billion was paid to Medicare beneficiaries for in-hospital costs when CHD was the principal diagnosis ($14 009 per discharge for AMI, $12 977 per discharge for coronary atherosclerosis, and $10 630 per discharge for other ischemic HD). [5]

•On the basis of data from NHANES 2007 to 2010 (NHLBI tabulation), an estimated 15.4 million Americans =20 years of age have CHD.
  -Total CHD prevalence is 6.4% in US adults >=20 years of age. CHD prevalence is 7.9% for men and 5.1% for women.
  -Among non-Hispanic whites, CHD prevalence is 8.2% for men and 4.6% for women.
  -Among non-Hispanic blacks, CHD prevalence is 6.8% for men and 7.1% for women
  -Among Mexican Americans, CHD prevalence is 6.7% for men and 5.3% for women.

•On the basis of data from the 2012 NHIS
  -Among Hispanic or Latino individuals >=18 years of age, CHD prevalence is 5.9%.1
  -Among American Indian/Alaska Natives >=18 years of age, it is estimated that 8.1% have CHD, and among Asians >=18 years of age, the estimate is 4.5%[6]

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

[1] Berry JD, Dyer A, Cai X, Garside DB, Ning H, Thomas A, Greenland P, Van Horn L, Tracy RP, Lloyd-Jones DM. Lifetime risks of cardiovascular disease. N Engl J Med. 2012;366:321–329.

[2] Roger VL, Go AS, Lloyd-Jones DM, et al; on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2014 update: a report from the American Heart Association. Circulation. 2014;129:e28–e292.

[3] Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, Finkelstein EA, Hong Y, Johnston SC, Khera A, Lloyd-Jones DM, Nelson SA, Nichol G, Orenstein D, Wilson PW, Woo YJ; on behalf of the American Heart Association Advocacy Coordinating Committee; Stroke Council; Council on Cardiovascular Radiology and Intervention; Council on Clinical Cardiology; Council on Epidemiology and Prevention; Council on Arteriosclerosis, Thrombosis and Vascular Biology; Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation; Council on Cardiovascular Nursing; Council on the Kidney in Cardiovascular Disease; Council on Cardiovascular Surgery and Anesthesia, and Interdisciplinary Council on Quality of Care and Outcomes Research. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association.

Circulation. 2011;123:933–944.

[4] Thom TJ, Kannel WB, Silbershatz H, D'Agostino RB Sr. Cardiovascular diseases in the United States and prevention approaches. In: Fuster V, Alexander RW, O'Rourke RA, Roberts R, King SB 3rd, Wellens HJJ, eds. Hurst's the Heart. 10th ed. New York, NY: McGraw-Hill; 2001:3–18.

[5] Foraker RE, Rose KM, McGinn AP, Suchindran CM, Goff DC Jr, Whitsel EA, Wood JL, Rosamond WD. Neighborhood income, health insurance, and prehospital delay for myocardial infarction: the atherosclerosis risk in communities study. Arch Intern Med. 2008;168:1874–1879.

[6] Blackwell D, Lucas J, Clarke T. Summary health statistics for U.S. adults: National Health Interview Survey, 2012. National Center for Health Statistics. Vital Health Stat 10.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not Applicable.

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 subcriteria 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 : Ischemic Heart Disease, Coronary Artery Disease

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

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.)*
http://www.ama-assn.org/ama1/pub/upload/mm/pcpi/cadminisetjune06.pdf

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
For endorsement maintenance, we are not including e-measure specifications as we did in the previous submission cycle. We have

updated the measure specifications to included are the applicable CPT and ICD-10 codes.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who were prescribed* aspirin or clopidogrel within a 12 month period.

*Prescribed may include prescription given to the patient for aspirin or clopidogrel at one or more visits in the measurement period OR patient already taking aspirin or clopidogrel as documented in current medication list.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Once during the measurement period.

S.6. **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, 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.
For Claims/Administrative: Report CPT II Code 4086F: Aspirin or clopidogrel prescribed.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
See 'Registry Supplemental Resources' attached in appendix field A.1 for data dictionary and form.

Codes that are applicable for the denominator are:

Diagnosis for coronary artery disease (ICD-9-CM) 410.00, 410.01, 410.02, 410.10, 410.11, 410.12, 410.20, 410.21, 410.22, 410.30, 410.31, 410.32, 410.40, 410.41, 410.42, 410.50, 410.51, 410.52, 410.60, 410.61, 410.62, 410.70, 410.71, 410.72, 410.80, 410.81, 410.82, 410.90, 410.91, 410.92, 411.0, 411.1, 411.81, 411.89, 412, 413.0, 413.1, 413.9, 414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.2, 414.3, 414.8, 414.9, V45.81, V45.82

Diagnosis for coronary artery disease (ICD-10-CM): I20.0, I20.1, I20.8, I20.9, I21.01, I21.02, I21.09, I21.11, I21.19, I21.21, I21.29, I21.3, I21.4, I22.0, I22.1, I22.2, I22.8, I22.9, I24.0, I24.1, I24.8, I24.9, I25.10, I25.110, I25.111, I25.118, I25.119, I25.2, I25.5, I25.6, I25.700, I25.701, I25.708, I25.709, I25.710, I25.711, I25.718, I25.719, I25.720, I25.721, I25.728, I25.729, I25.730, I25.731, I25.738, I25.739, I25.750, I25.751, I25.758, I25.759, I25.760, I25.761, I25.768, I25.769, I25.790, I25.791, I25.798, I25.799, I25.810, I25.811, I25.812, I25.82, I25.83, I25.89, I25.9, Z95.1, Z95.5, Z98.61

Patient encounter during the reporting period (CPT): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99304, 99305, 99306, 99307, 99308, 99309, 99310, 99324, 99325, 99326, 99327, 99328, 99334, 99335, 99336, 99337, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Documentation of medical reason(s) for not prescribing aspirin or clopidogrel (e.g., allergy, intolerance, receiving other

thienopyridine therapy, receiving warfarin therapy, bleeding coagulation disorders, other medical reasons)

Documentation of patient reason(s) for not prescribing aspirin or clopidogrel (e.g., patient declined, other patient reasons)

Documentation of system reason(s) for not prescribing aspirin or clopidogrel (e.g., lack of drug availability, other reasons attributable to the health care system)

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
For Claims/Administrative:

Documentation of medical reason(s) for not prescribing aspirin or clopidogrel
• Append modifier to CPT II code 4086F-1P

Documentation of patient reason(s) for not prescribing aspirin or clopidogrel
• Append modifier to CPT II code 4086F-2P

Documentation of system reason(s) for not prescribing aspirin or clopidogrel
• Append modifier to CPT II code 4086F-3P

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not Applicable.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not Applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not Applicable.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)
To calculate performance rates:

1) Find the patients who meet the initial patient 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 patient 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 patient population and denominator are identical.

3) Find the patients who quality for exclusions and subtract from the denominator.

4) From the patients within the denominator (after exclusions have been subtracted from the denominator), find the patients who qualify for the Numerator (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

5) From the patients who did not meet the numerator criteria, determine if the physician has documented that the patient meets any criteria for exception when exceptions have been specified [for this measure: medical reason(s)(e.g., eg, allergy, intolerance, receiving other thienopyridine therapy, receiving warfarin therapy, bleeding coagulation disorders, other medical reasons) or patient reason(s)(e.g., economic, social, and/or religious impediments, noncompliance, patient refusal, other patient reason)]. 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 of patients with valid exceptions) should be calculated and reported along with performance rates to track variations in care and highlight possible areas of focus for QI. If the patient does not meet the numerator and a valid exception is not present, this case represents a quality failure.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not Applicable. This measure is not based on sampling.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable. This measure is not based on a survey.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If data required to determine if an individual patient should be included in a specific performance measure based on defined criteria is missing, those cases would ineligible for inclusion in the denominator and therefore the case would be deleted. If data required to determine if a denominator eligible patient qualifies for the numerator (or has a valid exclusion/exception) is missing, this case would represent a quality failure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is currently being used in the ACCF PINNACLE registry for the outpatient office setting.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
CAD_Antiplatelet_0067_Testing_Form_Version_6.5_6.23.15.pdf

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0067
Measure Title: Chronic Stable Coronary Artery Disease: Antiplatelet Therapy
Date of Submission: 6/23/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record

☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The primary analysis was conducted at the level of the individual provider and included all patients with coronary artery disease (CAD) cared for by that provider and captured in the PINNACLE Registry during the one-year study period. The PINNACLE Registry systematically maps each practice's Electronic Health Record to the data elements required for the Registry, with careful validation of the translation process prior to enrollment. Data from the registry are reported back to the practices on a quarterly basis for quality improvement and are available for CMS reporting. Using these data, we were able to calculate the number of patients who should have received antiplatelet therapy, or a clinically, evidence-based reason not to use antiplatelet therapy was documented. This means that every patient in that provider's practice is included. For this measure, providers with less than 10 eligible patient encounters during the study period were excluded, since performance estimates are unstable with such small numbers. All other cases from all practices and providers were included. We included all visits for each patient in these analyses and meeting the performance measure on any single visit within the year met the criterion for this measure.

1.3. What are the dates of the data used in testing? The primary analysis included encounters between 01/01/2014-12/31/2014. Additionally, we used data from 01/01/2013 thru 12/31/2013 for temporal comparison.

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

<u>2013</u>

2,407 providers met the minimum number of eligible patients (10) for inclusion in the reliability analysis. The average number of eligible patients for providers included is 425.2 for a total of 1,023,530 patients. The range of number of patients for providers included is from 10 to 2,834. As described above, providers with fewer than 10 eligible patient encounters during the study period were excluded.

The unit of analysis for this measure is the provider. A description of the providers studied for the 2013 calendar year is shown below:

	Total **n = 2407**
Provider gender	
(1) Male	1923 (80.0%)
(2) Female	482 (20.0%)
Missing (.)	2
Provider categories	
NP/PA	258 (10.9%)
MD/DO	2061 (86.9%)
RN/nurses	52 (2.2%)
Missing (.)	36
Region	
(1) Northeast	305 (12.7%)
(2) Midwest	698 (29.0%)
(3) South	956 (39.7%)
(4) West	448 (18.6%)

2014

2,248 providers met the minimum number of eligible patients (10) for inclusion in the reliability analysis. The average number of eligible patients for providers included is 427.0 for a total of 959,792 patients. The range of numbers of patients for providers included is from 10 to 2,649. As described above, providers with fewer than 10 eligible patient encounters during the study period were excluded.

The unit of analysis for this measure is the provider. A description of the providers studied for the 2014 calendar year is shown below:

	Total **n = 2248**

	Total
	n = 2248
Provider gender	
(1) Male	1784 (79.4%)
(2) Female	464 (20.6%)
Provider categories	
NP/PA	250 (11.3%)
MD/DO	1915 (86.7%)
RN/nurses	44 (2.0%)
Missing (.)	39
Region	
(1) Northeast	257 (11.4%)
(2) Midwest	640 (28.5%)
(3) South	905 (40.3%)
(4) West	446 (19.8%)

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

<u>2013</u>

There are a total of 1,023,530 patients included in the temporal comparison that were treated in 2013. Patients' characteristics are provided below:

	Total
	n = 1023530
Race	
(1) White	502521 (92.0%)
(2) Black	31466 (5.8%)
(3) Other	12014 (2.2%)
Missing (.)	170072
Insurance	
(0) No insurance	1381 (1.0%)
(1) Private	100517 (69.9%)
(2) Medicare	40963 (28.5%)
(3) Medicaid	182 (0.1%)
(4) Other	820 (0.6%)
Missing (.)	154582
Age	
18 to <60	197507 (19.3%)
60 to <70	285197 (27.9%)
70 to <80	320245 (31.3%)
80 to 114	220581 (21.6%)

	Total n = 1023530
Sex (1) Male (2) Female Missing (.)	 637419 (62.4%) 384062 (37.6%) 1918
BMI (kg/m2) Missing	30.6 ± 9.0 231647
Diabetes Mellitus	288906 (28.9%)
Hypertension	807299 (85.4%)
Atrial Fibrillation/Flutter	217128 (22.7%)
Heart Failure	239885 (24.5%)
Peripheral Arterial Disease	139911 (16.1%)
Stroke/TIA	40229 (6.1%)
Myocardial Infarction	261043 (30.3%)

2014

There are a total of 959,792 patients included in the primary analysis (2014). See below for details on patient characteristics.

	Total n = 959792
Race (1) White (2) Black (3) Other Missing (.)	 428912 (92.9%) 23387 (5.1%) 9331 (2.0%) 158682
Insurance (0) No insurance (1) Private (2) Medicare (3) Medicaid (4) Other Missing (.)	 528 (0.4%) 99826 (81.6%) 21467 (17.6%) 177 (0.1%) 284 (0.2%) 158102
Age 18 to <60 60 to <70 70 to <80 80 to 114	 179194 (18.7%) 265376 (27.6%) 306202 (31.9%) 209020 (21.8%)
Sex (1) Male (2) Female Missing (.)	 599619 (62.6%) 357647 (37.4%) 2248

	Total
	n = 959792
BMI (kg/m2) Missing	30.7 ± 8.9 189212
Diabetes Mellitus	265083 (29.4%)
Hypertension	724628 (86.1%)
Atrial Fibrillation/Flutter	210886 (23.9%)
Heart Failure	233017 (26.1%)
Peripheral Arterial Disease	136158 (16.3%)
Stroke/TIA	40536 (6.2%)
Myocardial Infarction	238396 (29.3%)

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

The dataset described above was used for all aspects of testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We do not currently collect any of the SDS variables examples listed above. As is noted in other sections of this testing form we do collect data on race as well as insurance type.

2a2. RELIABILITY TESTING

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

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

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

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

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

Reliability 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], where the latter represents the within-physician estimate of our error in assessing their 'true' performance. Thus, the reliability estimate 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 five different distributions of providers' patient volumes: at the minimum number of quality reporting events for the measure; at the mean number of quality reporting events per physician; and at the 25th, 50th and 75th percentiles of the number of quality reporting events.

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*)

2013 – In 2013, the signal-noise ratios are shown below:

Description	Number of Patients	Signal-to-Noise Ratio
Minimum	10	0.994
25th percentile	188	0.998
50th percentile	368	0.998
75th percentile	586	0.999
Average	426	0.998

2014 – In 2014, the signal-noise ratios are shown below:

Description	Number of Patients	Signal-to-Noise Ratio
Minimum	10	0.995
25th percentile	204	0.998
50th percentile	376	0.998
75th percentile	578	0.999
Average	427	0.998

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?*)

For this measure the reliability was very high and was similar for 2013 and 2014, supporting the reproducibility of these estimates across years. At the minimum number of patient visits required (>10) the average reliability was 0.994 and 0.986 for 2013 and 2014, respectively. For providers with the median number of patient encounters, the reliability was even higher, with 0.998 for both years. Given that a reliability of 0.70 is generally considered a minimum threshold for acceptability, and 0.80 is considered very good reliability, these data suggest that the measure is exceedingly good at describing true differences across physicians.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

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

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. 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)*

Content validity for this measure was assessed by expert work group members during the development process. Additional input on the content validity of draft measures was established through a 30-day public comment period and concurrent formal peer review process. All comments received were reviewed by the expert work group and the measures were adjusted as needed. Additionally, the measure underwent review and approval by the Board of Trustees of the ACC and the Science Advisory and Coordinating Committee of the AHA, as well as review and voting by the PCPI membership. Members of the expert work group that developed the measure included: Bruce Abramowitz, MD, FACC (interventional cardiology; measure implementation), Karen Alexander, MD (cardiology; geriatrics), Craig T. Beam, CRE (patient representative), Robert O. Bonow, MD, MACC, FAHA, FACP (cardiology), Jill S. Burkiewicz, PharmD, BCPS (pharmacy); Michael Crouch, MD, MSPH (family medicine), David C. Goff, Jr., MD, PhD, FAHA, FACP (internal medicine), Richard Hellman, MD, FACP, FACE (endocrinology), Thomas James, III, FACP, FAAP (health plan representative), Marjorie L. King, MD, FACC, MAACVPR (cardiology; cardiac rehabilitation), Edison A. Machado, Jr., MD, MBA (measure implementation), Eduardo Ortiz, MD, MPH (guideline development), Michael O'Toole, MD (cardiology; electrophysiology; measure implementation), Stephen D. Persell, MD, MPH (internal medicine; measure implementation), Jesse M. Pines, MD, MBA, MSCE, FAAEM (emergency medicine), Frank J. Rybicki, MD, PhD (radiology), Lawrence B. Sadwin (patient representative), Joanna D. Sikkema, MSN, ANP-BC, FAHA (cardiology), Peter K. Smith, MD (thoracic surgery), Patrick J. Torcson, MD, FACP, MMM (hospital medicine), John B. Wong MD, FACP (internal medicine).

Construct validity was difficult to establish because there has not been an independent audit of these data. However, it is important to note that an independent audit would merely involve an abstractor reviewing the

same medical record from which PINNACLE directly abstracts its data and, given the identical source of the data, any error observed would either be due to the auditor incorrectly abstracting the data from the EHR or PINNACLE incorrectly mapping the data elements from the EHR. To address the latter, we conduct detailed analyses to insure that this does not happen and quarantine (i.e. not report) data that fails our addition Data Quality Review process. Validity of measure data elements in PINNACLE is routinely evaluated on a quarterly basis as part of the standard data extraction and analytic data set creation process. First, all relevant data elements are reviewed at the record level to ensure that individual data values are valid; any invalid values are set to missing. Next, the *distribution* of each data element is reviewed, aggregating both across practices and across calendar quarters within each practice, to identify outliers, suspicious patterns and/or systematic changes in the prevalence of the data element that may suggest data mapping errors or unanticipated changes in definitions, coding consistency, data completeness, etc. Identification of suspicious data includes both statistical criteria, using quality control charts with rigorous definitions of "out of control" rates, and manual clinical review of each distribution for plausibility. Records that are flagged as suspicious by these criteria are quarantined and excluded from analysis and reporting. In 2013 the rate of records not passing the quality evaluation was 1.9% and in 2014 it was 2.1%. Feedback reports are generated to facilitate investigation of data issues at the practice level to verify accuracy of abstraction and to remap elements whose definitions or recording have changed.

Face validity of the measure score was systematically assessed as follows:

After the measure was fully specified, members of two existing committees, one at the ACC and one at AHA, with expertise in in general cardiology, interventional cardiology, heart failure, electrophysiology and quality improvement, outcomes research, informatics and performance measurement, who were not involved in development of the measure, were asked to review the measure specifications and 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."

The respondents recorded their rating on a scale of 1-5, where 1= Strongly Disagree; 3=Neither Agree nor Disagree; 5= Strongly Agree

Forty Two (42) members completed the survey and provided a mean importance rating of 4.26, with 83.3% agreeing with the use of the measure for quality assessment.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

We believe that the processes used to extract data from the exact source from which any abstraction process done manually would use (the EHR), and our thorough data quality review, provide strong evidence for the validity of this measure.

The results of the expert panel rating of the validity statement were as follows:

N = 42; Mean rating = 4.26 and 83.3% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings
1 - <2> (Strongly Disagree)
2 - <3>
3 - <2> (Neither Agree nor Disagree)
4 - <10>
5 - <25> (Strongly Agree)

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The measure was judged to have high face validity by both its clinical importance and by the group of experts asked to rate it. The majority of experts agreed that the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality. Importantly, as a process measure, the strong association of treatment with improved survival and reduced myocardial infarction rates provide strong validity for this measure as a mechanism to insure that strong clinical evidence is being translated to routine clinical care.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Since not all patients with CAD will meet the guideline recommendations for antiplatelet therapy, exclusions in this measure are intended to remove patients for whom antiplatelet therapy may not be appropriate. We divide these into two categories: Exclusions and Exceptions. Exclusions arise when patients who are included in the initial patient or eligible population for the measure set do not meet the denominator criteria specific to the intervention required by the numerator. Exclusions are absolute and apply to all patients and therefore are not part of clinical judgment within a measure. Specific exclusions should be derived from evidence-based guidelines. Exceptions are not absolute, and are based on clinical judgment and individual patient characteristics, thus patients with such contraindications represent circumstances where the clinicians balanced the risks and benefits and felt that, in a given situation, the benefits outweighed the risks and chose to treat the patient. These patients are therefore included in both the numerator and denominator of the measure. In contrast, the exceptions are clearly documented reasons to not treat the patient and are removed from the denominator of the population.

Exclusions in this measure:

- Documented medical, patient, or system reason for not on aspirin or clopidogrel or patients were on warfarin.

2b3.2. 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*)

The Exceptions for each year are provided below:

2013: 3.1% (n=74) of the providers do not have exceptions in the denominator. Among 2,333 providers who do have exceptions, the exclusion rate ranges from 0.2% to 35.3%, mean is 6.0%. Of those patients removed from the measure, 0.72% were removed due to a medical reason, 9.47% were removed for a patient reason, 1.48% were removed for a system reason, 0.30% were removed due to multiple reasons, and 88.03% were removed due to patients who were on warfarin or another thienopyridine that is not included in this measure (e.g. prasugrel or ticagrelor) and the physician felt that the addition of aspirin or a clopidogrel provided an increased risk of bleeding with minimal benefits in the prevention of recurrent cardiovascular events.

2014: 2.5% (n=56) of the providers do not have exceptions in the denominator. Among 2,192 providers who do have exceptions, the exclusion rate ranges from 0.3% to 27.4%, mean is 5.8%. Of those patients removed from the measure, 0.80% were removed due to a medical reason, 9.52% were removed for a patient reason, 1.88% were removed for a system reason, 0.31% were removed due to multiple reasons, and 87.50% were removed due to concomitant medications that either provide anti-platelet or systemic anti-coagulation.

2b3.3. 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*)

We do not view any concerns with exclusions or exceptions for this measure. While the majority of exceptions are due to patient reasons and these might be 'gameable' by clinicians, we recognize the bleeding and bruising are very frequent side effects of anti-platelet treatment (see Amin et al, *J Am Coll Cardiol* 2013; 2013;61(21):2130-8) and that clinicians who think enough about this decision process to document an exclusion are likely providing excellent, patient-centered care.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

We examined variation in provider performance on this measure based on sex, age, race and a number of other patient factors to identify variations. The findings are represented for 2013 and 2014 respectively.

2013

label	# of providers	# of patients	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
Male	2403	637419	0.00%	85.7%	88.6%	94.7%	100%	9.02%	9.97%
Female	2403	384062	0.00%	77.1%	82.5%	90.8%	100%	13.8%	12.4%
Age: <60	2404	197507	0.00%	73.9%	81.6%	92.7%	100%	18.7%	15.1%
Age: 60 -< 70	2406	285197	0.00%	83.9%	87.5%	94.5%	100%	10.6%	11.1%
Age: 70 -< 80	2405	320245	0.00%	84.6%	87.9%	94.4%	100%	9.80%	10.6%
Age: >= 80	2400	220581	0.00%	83.0%	87.1%	94.4%	100%	11.4%	11.3%
Insurance: None	134	1381	0.00%	92.3%	86.1%	100%	100%	7.69%	28.6%
Insurance: Private	683	100517	0.00%	83.9%	87.5%	96.0%	100%	12.1%	13.8%
Insurance: Medicaid	500	40963	0.00%	84.7%	89.5%	100%	100%	15.3%	13.1%
Insurance: Medicare	27	182	50.0%	86.7%	92.3%	100%	100%	13.3%	12.5%
Insurance: Other	152	820	0.00%	68.3%	81.4%	100%	100%	31.7%	29.2%
Race: White	1552	502521	0.00%	83.7%	86.8%	93.7%	100%	10.1%	10.9%
Race: Black	1425	31466	0.00%	77.8%	83.9%	100%	100%	22.2%	19.9%
Race: Other	1190	12014	0.00%	75.0%	83.8%	100%	100%	25.0%	25.6%

2014

Label	# of providers	# of patients	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
Male	2243	599619	0.00%	85.7%	88.8%	94.5%	100%	8.80%	8.84%
Female	2242	357647	0.00%	76.6%	82.2%	90.6%	100%	14.0%	11.7%
Age: <60	2246	179194	0.00%	74.7%	82.0%	92.1%	100%	17.3%	14.2%
Age: 60 -< 70	2248	265376	0.00%	83.7%	87.5%	94.1%	100%	10.5%	10.0%
Age: 70 -< 80	2248	306202	20.0%	83.9%	87.9%	94.3%	100%	10.4%	9.50%
Age: >= 80	2242	209020	0.00%	83.0%	87.1%	93.9%	100%	10.9%	10.5%
Insurance: None	104	528	0.00%	100%	90.5%	100%	100%	0.00%	24.1%
Insurance: Private	606	99826	0.00%	84.3%	88.6%	96.5%	100%	12.1%	12.2%
Insurance: Medicaid	397	21467	0.00%	85.0%	89.8%	100%	100%	15.0%	13.4%
Insurance: Medicare	27	177	0.00%	80.0%	87.9%	100%	100%	20.0%	21.5%
Insurance: Other	103	284	0.00%	77.8%	83.0%	100%	100%	22.2%	31.8%
Race: White	1288	428912	0.00%	83.7%	87.1%	93.3%	100%	9.58%	9.79%
Race: Black	1212	23387	0.00%	77.2%	84.7%	100%	100%	22.8%	18.4%
Race: Other	1090	9331	0.00%	76.9%	84.6%	100%	100%	23.1%	25.4%

2b5.2. 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*)

2013

Overall mean performance on this measure is 86.2%, with a standard deviation of 10.5%. The minimum score equals 0.00%, while the maximum score equals 100.00%. The interquartile score is equal to 10.3%.

2,407 providers were measured, and the patient study sample equals 1,023,530. 62.4% of the sample is male. 92.0% of the sample is white, 5.8% is black, and 2.2 % identified as "other." The sample reached across all US regions, with 12.7% of providers in the Northeast, 29.0% of providers in the Midwest, 39.7% of providers in the South, and 18.6% of providers in the West.

# of providers	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
2407	0.00%	82.6%	86.2%	92.9%	100%	10.3%	10.5%

	Mean
Decile 1	62.6%
Decile 2	77.7%
Decile 3	82.6%
Decile 4	85.5%
Decile 5	87.6%
Decile 6	89.5%
Decile 7	91.2%
Decile 8	93.0%
Decile 9	94.9%
Decile 10	97.6%

2014

Overall mean performance on this measure is 86.3%, with a standard deviation of 9.49%. The minimum score equals 0.00%, while the maximum score equals 100.00%. The interquartile score is equal to 10.2%.

2,248 providers were measured, and the patient study sample equals 959,792. 62.6% of the sample is male. 92.9% of the sample is white, 5.1% is black, and 2.0% identified as "other." The sample reached across all US regions, with 11.4% of providers in the Northeast, 28.5% of providers in the Midwest, 40.3% of providers in the South, and 19.8% of providers in the West.

# of providers	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
2248	10.0%	82.6%	86.3%	92.8%	100%	10.2%	9.49%

	Mean
Decile 1	65.4%
Decile 2	77.8%
Decile 3	82.4%
Decile 4	85.1%
Decile 5	87.3%
Decile 6	89.2%
Decile 7	91.0%
Decile 8	92.8%
Decile 9	94.7%
Decile 10	97.4%

2b5.3. 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?*)

2013: A small amount of variability was noted among providers. The performance-met rate range was 0-100% with the inter-quartile range being 82.6% to 92.9%. This yielded a Median Rate Ratio of 1.11 (1.10, 1.11). The Median Rate Ratio measures the variation across providers for statistically 'identical' patients and suggests that a patient presenting to 1 provider, as opposed to another, would, on average, be 11% more likely to be treated with an anti-platelet agent.

2014: A small amount of variability was also noted among providers in 2014. The performance-met rate range was 10-100% with the inter-quartile range being 83% to 93%. This yielded a Median Rate Ratio of 1.09 (1.09, 1.10).

2b6. 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 SDS 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 SDS 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.

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

In PINNACLE missing values are interpreted as "no" for most variables. It is challenging to distinguish real missing versus "No." Therefore, we are assuming that missing documentation of antiplatelet indicates a failure to meet the measure. It is possible that a provider may not have documented antiplatelet therapy use in their EMR system, perhaps if it was provided by another provider in a different healthcare system. However, we believe that this is unlikely and that it is still incumbent upon a provider to clearly document all the medications that a patient is taking, particularly antiplatelet for CAD patients.

2b7.2. 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*)

Given our assumptions, noted above, we did not conduct an empirical analysis of frequency or distribution of missing data. For this measure, missing data is reported as a quality failure.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

We do not believe any biases are introduced in the assessing of individual physician performance and continued endorsement of this measure would lead to improved care.

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)

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)*

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

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).

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)*.

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, the ACC, the AHA, the PCPI and its members disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT®) or other coding contained in the specifications. CPT® contained in the measures specifications is copyright 2004-2015 American Medical Association.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement (Internal to the specific organization)	Payment Program Physician Quality Reporting System http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment- Regulatory and Accreditation Programs Physician Quality Reporting System http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment- Quality Improvement with Benchmarking (external benchmarking to multiple organizations) ACC Pinnacle Registry URL: http://www.ncdr.com/webncdr/pinnacle/

4a.1. For each CURRENT use, checked above, provide:

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

PINNACLE Registry (URL: http://www.ncdr.com/webncdr/pinnacle/)
In 2008, the American College of Cardiology Foundation launched the PINNACLE program (Formerly known as the Improving Continuous Cardiac Care or IC3). The PINNACLE Registry® continues to grow rapidly, with more than 3446 providers representing almost 960 unique office locations across the U.S submitting data to the registry. As of March 2015, the registry has more than 19.8 million patient encounter records representing approximately 4.85 million patients. PINNACLE assists practices in understanding and improving care through the production and distribution of quarterly performance reports. These reports, covering all valid patient encounters, detail adherence to 28 cardiovascular clinical measures at the physician, location, and practice levels across coronary artery disease, hypertension, heart failure and atrial fibrillation.

PQRS Qualified Clinical Data Registry:
In addition to the current use for quality improvement with benchmarking in the PINNACLE registry, the measure will be reported to CMS by the registry as part of PQRS in 2015 as a Qualified Clinical Data Registry. Pinnacle did submit data for 2014 for QCDR purposes to CMS. Eligible professionals will be considered to have satisfactorily participated in PQRS if they submit quality measures data or results to CMS via a qualified clinical data registry. CMS does have plans to eventually publicly report QCDR data.

CAD: Antiplatelet Therapy
Physician Quality Reporting System (Centers for Medicare and Medicaid Services):
PQRS is a national reporting program that uses a combination of incentive payments and payment adjustments to promote

reporting of quality information by eligible professionals (EPs). The program provides an incentive payment to practices with Eps (identified on claims by their individual National Provider Identifier [NPI] and Tax Identification Number [TIN]). EPs satisfactorily report data on quality measures for covered Physician Fee Schedule (PFS) services furnished to Medicare Part B Fee-for-Service (FFS) beneficiaries (including Railroad Retirement Board and Medicare Secondary Payer). Beginning in 2015, the program also applies a payment adjustment to EPs who do not satisfactorily report data on quality measures for covered professional services. Source: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html It is our understanding that CMS is also planning to move towards publicly reporting physician data via Physician Compare.

4a.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 are continuously seeking opportunities to advocate for 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.

4a.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.

Also, although the measure is currently in use, we are continuously seeking opportunities to advocate for expanded use of this measure in government or other programs, including those intended for accountability or public reporting.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

The mean performance rates from the Pinnacle registry was 86.2% in 2013 and 86.3% in 2014.

In 2013, 2,407 providers were measured, and the patient study sample equals 1,023,530. In 2014, 2,248 providers were measured, and the patient study sample equaled 959,792. The statistical significance of these results was not analyzed.

While the ACC and AHA 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.

4b.2. **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.**

Not Applicable.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

We are not aware of any unintended consequences at this time, but we continuously monitor for them.

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)**

0465 : Perioperative Anti-platelet Therapy for Patients undergoing Carotid Endarterectomy

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

5a. Harmonization

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 completely harmonized?

No

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

See 5b.1 for more detailed response due to lack of character spaces in this section.

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.)

Measure 0067 looks at whether ASA or clopidogrel where prescribed during a 12 month measurement period. Meanwhile, the two

existing NQF endorsed measures (#0465 and #0964) focused on whether the medications were prescribed prior to discharge or prior to surgery.

Specifically, Measure #0465 (Perioperative Antiplatelet Therapy for patients undergoing Carotid Endaroretomy) focuses on inpatient who were provided ASA or clopidogrel within 48 hours prior to surgery and prescribed this medication at hospital discharge. Measure #0067 looks at whether ASA or clopidogrel was prescribed during the 12 month measurement period. Both measures allow for medical exceptions.

In the case of Measure 0964 (Therapy with aspirin, P2Y12 inhibitor, and statin at discharge following PCI in eligible patients), this measure is also an inpatient measure and focuses on sosley PCI eligible patients who had ASA or P2y12 and statins prescribed prior to discharge. Measure 0067 looks at whether ASA or clopidogrel was prescribed during the 12 month measurement period. Both measures allow for medical exceptions.

Measures #0465 and #0964 address a different patient demographic and focuses on inpatient prescribed of ASA or Clopidogrel.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** CAD_Antiplatelet_0067_PINNACLE_Registry_data_collection_form_and_dictionary-635705651789969036.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American College of Cardiology
Co.2 **Point of Contact:** Penelope, Solis, comment@acc.org, 202-375-6576-
Co.3 **Measure Developer if different from Measure Steward:** American College of Cardiology
Co.4 **Point of Contact:** Penelope, Solis, comment@acc.org, 202-375-6576-

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.
Bruce Abramowitz, MD, FACC (interventional cardiology; measure implementation)
Karen Alexander, MD (cardiology; geriatrics)
Craig T. Beam, CRE (patient representative)
Robert O. Bonow, MD, MACC, FAHA, FACP (cardiology)
Jill S. Burkiewicz, PharmD, BCPS (pharmacy)
Michael Crouch, MD, MSPH (family medicine)
David C. Goff, Jr., MD, PhD, FAHA, FACP (internal medicine)
Richard Hellman, MD, FACP, FACE (endocrinology)
Thomas James, III, FACP, FAAP (health plan representative)
Marjorie L. King, MD, FACC, MAACVPR (cardiology; cardiac rehabilitation)
Edison A. Machado, Jr., MD, MBA (measure implementation)
Eduardo Ortiz, MD, MPH (guideline development)
Michael O'Toole, MD (cardiology; electrophysiology; measure implementation)
Stephen D. Persell, MD, MPH (internal medicine; measure implementation)
Jesse M. Pines, MD, MBA, MSCE, FAAEM (emergency medicine)
Frank J. Rybicki, MD, PhD (radiology)

Lawrence B. Sadwin (patient representative)
Joanna D. Sikkema, MSN, ANP-BC, FAHA (cardiology)
Peter K. Smith, MD (thoracic surgery)
Patrick J. Torcson, MD, FACP, MMM (hospital medicine)
John B. Wong MD, FACP (internal medicine)

ACC/AHA/ AMA PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study must be equal contributors to the measure development process. In addition, the ACC/AHA/AMA PCPI strive to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

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:** 01, 2011
Ad.4 **What is your frequency for review/update of this measure?** On average, every 3 years or as new evidence becomes available that materially affects the measures.
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.7 **Disclaimers:** This Physician Performance Measurement Set (PPMS) and related data specifications were developed by the Physician Consortium for Performance Improvement (the Consortium) including the American College of Cardiology (ACC), the American Heart Association (AHA) and the American Medical Association (AMA) to facilitate quality improvement activities by physicians. The performance measures contained in this PPMS are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. This PPMS is intended to assist physicians to enhance quality of care and is not intended for comparing individual physicians to each other or for individual physician accountability by comparing physician performance against the measure or guideline.

Ad.8 **Additional Information/Comments:** The ACCF, AHA, and PCPI have a formal measurement review process that stipulates regular (usually on a three-year cycle, when feasible) review of the measures. The process can also be activated if there is a major change in scientific evidence, results from testing or other implementation issues are noted that materially affect the integrity of the measure.

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 #: 0068
Measure Title: Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
Measure Steward: National Committee for Quality Assurance
Brief Description of Measure: The percentage of patients 18 years of age and older who were discharged from an inpatient setting with an acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary intervention (PCI) during the 12 months prior to the measurement year, or who had a diagnosis of ischemic vascular disease (IVD) during the measurement year and the year prior to the measurement year and who had documentation of routine use of aspirin or another antiplatelet during the measurement year.
Developer Rationale: This measure aims to improve the quality of care for patients with established cardiovascular disease by assessing whether or not they received aspirin or another antiplatelet medication during the measurement year. Antiplatelet medications, such as aspirin and clopidogrel, are drugs that inhibit platelets from clumping together and forming clots. Their use in the secondary prevention of cardiovascular events is well-established. In patients who are at high risk because they already have atherosclerotic cardiovascular disease, antiplatelet therapy reduces the yearly risk of serious vascular events (MI, stroke, death) by about twenty-five percent. Antiplatelet agents also have a beneficial effect in reducing all-cause mortality and fatal cardiovascular events in patients with peripheral arterial disease.

Numerator Statement: Patients who had documentation of routine use of aspirin or another antiplatelet during the measurement year.
Denominator Statement: Patients 18 years or older by the end of the measurement year discharged from an inpatient setting with an AMI, CABG, or PCI during the 12 months prior to the measurement year or who had a diagnosis of IVD during both the measurement year and the year prior to the measurement year.
Denominator Exclusions: Patients who had documentation of use of anticoagulant medications during the measurement year.

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records
Level of Analysis: Clinician : Group/Practice, Clinician : Individual
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

The measure is included in CMS' Electronic Health Record Incentive Program Stage 2 and ACO Shared Savings Program. An eMeasure was submitted for endorsement review.

Is this a MAINTENANCE measure submission? ☒ Yes ☐ No, this is a NEW measure submission.
For a MAINTENANCE, what is the Original Endorsement Date: 8/10/09 Most Recent Endorsement Date: 1/18/12

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

Public & Member Comments included:
• Competing measures contain differences with respect to data collection methods, applicable settings, and exclusion criteria; however, it's important that the Steering Committee continue to work with developers of measures #0068, #0067, #0075 to determine the feasibility of harmonizing specifications of these measures where appropriate.
• Favor composite measure 0076 over the individual measures.
• Add BRILINTA (ticagrelor) to the list of oral anti-platelet agents.

• Encourage the measure developer to commit to develop an all-or-nothing composite for its IVD process measures in the near term.
Developer Response:
• Inclusion of Brilinta will be reviewed during our routine measure update process which includes review by our pharmacy panel.
Steering Committee:
• Urged the developers to work toward harmonization of the measures.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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.

The developer provides the following evidence for this process measure:

- This is a clinician level measure that calculates the % of patients 18 years of age and older who were discharged from an inpatient setting with an acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary intervention (PCI) during the 12 months prior to the measurement year, or who had a diagnosis of ischemic vascular disease (IVD) during the measurement year and the year prior to the measurement year and who had documentation of routine use of aspirin or another antiplatelet during the measurement year in a clinician office/clinic.
- The developer provides 4 separate guidelines with 10 guideline statements for the use of aspirin or another antiplatelet in patients diagnosed with cardiovascular disease but does not provide the QQC for the clinical guidelines. Per NQF criteria, they reference one seminal systematic review cited in the guidelines that summarizes the body of evidence supporting the recommendations.
- The developer provides decision logic from secondary prevention to outcome for the use of aspirin or another antiplatelet medication in reducing the risk of having an AMI, stroke, vascular complication or dying in patients with cardiovascular disease and peripheral artery disease.

Questions for the Committee:

- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- The developer provides data from NCQA's Heart/Stroke Recognition Program and CMS' PQRS:

Performance Data from NCQA's Heart/Stroke Recognition Program (HSRP) – Individual Clinician Level

YEAR	N	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2012	755	65%	24%	38%	44%	59%	89%	97%	45%
2013	659	84%	20%	50%	83%	91%	97%	100%	14%

CMS PHYSICIAN QUALITY REPORTING SYSTEM (PQRS)

YEAR	N	MEAN
2010	4491	75%
2011	8167	80%
2012	18291	55%
2013	17935	64%

- The developer does not provide sample characteristics or volumes. No data on disparities from the measure is provided, though the prevalence of heart disease, stroke, and peripheral arterial disease (PAD) based on sex, age, and race are provided. The developer also cites a study that found blacks are less likely than whites to use aspirin (29% vs. 44%, p = 0.008). Another study found that PAD patients with lower SES are less likely to receive antiplatelet and statin therapy.

- ***Questions for the Committee:***
  - *The PQRS performance declines significantly from 2011 and 2012, while volumes significantly increased? Did a change in the measure or program impact the performance? Similar declines were noted in 2012 to 2013 in HSRP*
  - *Is there a gap in care that warrants a national performance measure?*
  - *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - High rating for evidence. QQC provided, and high quality for guidelines and systematic review. Evidence applies directly to process.
   - As this is a process measure it is required to have a systematic review. This measure has clinical guidelines with supportive statements with 1A strength of the statements in the guideline, as well as a systematic review of 287 studies including 135,000 patients. I feel this measure has sufficient evidence to support the measure focus.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***
- High rating for evidence. QQC provided, and high quality for guidelines and systematic review. Evidence applies directly to process.
- The evidence relates directly to the process measured which is routine use of aspirin or another antiplatelet. Recommend rating as HIGH

1b. ***Committee's Comments on Performance Gap:***
- Some performance gap noted, though data confusing with poorer performance when volume increased. No disparities related to measure noted, but a couple studies note both racial and SES disparities.
- Concerns exist in CMS PQRS that the N is increasing but the mean % decreases from 2010 through 2013. The NCQA heart stroke recognition program individual clinician level demonstrates increase in % reporting, but decrease in N reporting.
- There is no performance data on disparities, but it appears according to evidence that there are disparities, including black vs white and lower SES. In addition, specific diseases may not be receiving appropriate care (e.g. PAD)

1c. ***Committee's Comments on Composite Performance Measure:***
- Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- Data sources to calculate performance include administrative claims, electronic clinical data, EHR, and paper medical records. ICD-9, ICD-10, CPT, HCPCS, and UBREV (billing codes). An ICD-9 to ICD-10 conversion methodology is not provided.
- Eligible patients identified by either event or diagnosis. If identified by event (MI, CABG, PCI), the event must occur during the year prior to the measurement year. However, if identified by diagnosis of IVD then must have had diagnosis during both the measurement year and the year prior to the measurement year.
- At a minimum 'documentation' is defined as a note in the MR indicating dates when aspirin or other antiplatelet was prescribed or documentation of prescription from another treating physician.
- Developer states since last maintenance the following changes to the measure have been made:
    - Changed "antithrombotic" to "antiplatelet" throughout the measure specification to more closely align with measure intent, which is to assess the use of aspirin or other antiplatelet medications in patients with ischemic vascular disease.
    - Added exclusion for patients on anticoagulant therapy based on feedback from measurement advisory panels.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- Reliability testing for this measure was conducted at the level of the performance measure score, using the data source and level of analysis specified. The primary analysis was conducted at the level of the individual provider and included 659 clinicians (35 eligible patients per clinician; total = 23,065 patients) who participated in NCQA's Heart/Stroke Program (HSRP) in 2013.
- The sampling methodology for reliability testing included provider established abstraction start date, and selecting 35 consecutive regressive patients who meet age, IVD diagnostic, patient duration with provider, and timing parameters for face-to-face visits.
- A signal-to-noise analysis using the beta-binomial model was conducted. This type of analysis, which is appropriate for the measure, quantifies the amount of variation in performance that is due to differences between providers (as opposed to differences that are due to random measurement error). The method results in a reliability statistic for each clinician.
- Overall reliability for this measure based on data submitted by 659 clinicians is 0.88. A reliability of 0.70 is generally considered a minimum threshold for acceptability, and 0.80 is considered very good reliability.

Questions for the Committee:

- *Specific questions on the method and results of reliability testing.*

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The clinical practice guidelines and body of evidence supporting this measure recommend the use of aspirin or other antiplatelets as secondary preventive treatment in patients with established cardiovascular disease.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- Systematic assessment of face validity consists of a four step standardized process called the HEDIS measure life cycle. The steps include identifying areas of interest/gaps in care, a literature review, measure development, field testing, public comment, and ongoing evaluation.
- The developers state this measure was tested for face validity with two expert panels, Cardiovascular Measurement Advisory Panel (8 physicians, 1 nurse) and NCQA's Clinical Programs Committee (17 members).
- The developer does not provide statistical results from validity testing but does state that the multiple NCQA panels agreed that the measure, as specified, accurately differentiates quality across clinicians.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer states that prior to 2015 the measure did not have any exclusions. However, developer stated since last maintenance added exclusion for patients on anticoagulant therapy based on feedback from measurement advisory.
- No statistical testing results on the impact of adding the exclusion of patients on anticoagulant therapy provided.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This process measure is not risk adjusted or stratified.

2b5. Meaningful difference:

- A box plot with 2012 and 2013 data is included with # of clinicians, average, lowest/highest rate but the developer only provides SD, percentiles, IQR and interpretation of the 2013 data.
- The overall mean performance on this measure is 84% (SD = 20%) for 2013.
- There is a statistically significant difference (14%) in performance between clinicians in the 25th and 75th percentiles. The gap is significantly higher (50%) between the 10th and 90th percentiles.
- The developers expect clinicians outside of the Heart/Stroke Recognition Program (HSRP) to show wider variation in performance.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- This measure submission includes the claims and registry submission for the measure, and not the eMeasure specifications for the EHR Incentive Program and ACO Shared Savings Programs. Therefore, not comparability of data sources is available.

Question for the Committee:

Does performance across data sources for non-eMeasures and eMeasures of the same measure differ?

2b7. Missing Data

- The developer states that this measure is collected with a complete sample, therefore no missing data on this measure.

Question for the Committee:

- *Does the developer have data demonstrating how many patients were excluded due to missing data?*

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- Measure specifications are clearly defined and consistent with the evidence. Measure likely to be implemented consistently.
- The data elements are clearly defined and the evidence clearly supports this measure.

2a2.: ***Committee's Comments on Reliability-Testing:***

- Reliability testing was done through beta-binomial model measuring signal-to-noise ratio, and the results demonstrated high reliability analysis. High rating.
- The developers have done an excellent job of revising the numerator and the denominator by changing anitcoagulant in numerator to antiplatelet, and have added very appropriate and thorough exclusions that are consistent with the evidence. One concern is the number of patients who experience atrial fibrillation and are placed on anticoagulants for this who may be post PCI, CABG, MI... they would all be excluded from this measure?
- The exclusions are of sufficient frequency across providers.
- The reliability score (.88) is high suggesting sufficient signal strength to discriminate performance between entities

2b1.: ***Committee's Comments on Validity-Specifications***:

- Measure specifications are clearly defined and consistent with the evidence. Measure likely to be implemented consistently.
- None

2b2.: ***Committee's Comments on Validity-Testing:***

- Face validity done, and results indicate a majority percentage believe measure can accurately distinguish good and poor quality. No statistical results of validity testing provided. Overall moderate validity.
- Empirical validity testing was not performed, but face validity of performance measure score was performed with 2 expert panels: one with 9 members (8 physicians, one nurse) and one with 17 members.
- This is not the optimal method of validity testing.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Exceptions noted. No apparent threats, but meaningful differences are noted in clinician performance. No risk adjustment done.
- According to the box plot provided by the developers of the heart stroke recognition program from 2012 and 2013 there is a 14% gap between first 1/4 and 3/4 percentiles but a 50% gap between 10 and 90%...indicating meaningful differences in performance. This indicates clinicians outside of HSRP program would probably demonstrate a larger variation.

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- The data source includes administrative claims, electronic clinical data, EHR, and paper medical records and readily available data occurring during patient care.
- The developer states that some data elements are in electronic sources but to allow for widespread reporting across health care practices, this measure is collected through multiple data sources including paper medical records.
- The developers do not provide specific information on the operational use of the measure but do state that clinicians participating in NCQA's Heart/Stroke Recognition Program have no difficulty collecting, interpreting, and reporting this measure. Additionally, the developers have received positive feedback on the use of this measure in the PQRS program.
- The developers encourage broad public use and dissemination of the measure and agreed that noncommercial uses do not require the consent. Practice use is not considered commercial use.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- Data collection obtained through administrative claims, EHRs, and other sources. No data collection barriers identified. No assessments of costs or abstraction time done. Moderate to high feasibility.
- The data elements are routinely generated and used during care delivery and could easily be captured in electronic form.

Criterion 4: Usability and Use

4. Usability and 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.

- The measure is publicly reported in PQRS; it is also used in CMS' EHR Incentive Program (Meaningful Use), ACO Shared Savings Program, and NCQA's Heart/Stroke Recognition Program (HSRP).

- In 2014 the Measure Applications Partnership (MAP) Clinician Workgroup supported this measure for the Physician Compare and Value-Based Payment Modifier Programs because the measure promotes alignment across programs, settings, and public- and private-sector efforts. Included in a MAP family of measures. Address program goals/requirements. The measure was previously supported by Clinician Workgroup for inclusion in Physician Compare and VBPM for clinician group reporting.

Questions for the Committee (as appropriate) *:*

- *Is the measure publicly reported?*
- *For maintenance measures – is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- Measure is currently implemented by CMS in PQRS, EHR Incentive Program, and Medicare Shared Savings Program. High usability.
- Currently used in PQRS with CMS and in NCQAs heart stroke recognition program

Criterion 5: Related and Competing Measures

- List any related or competing measures based on harmonization protocol.
- Summarize any harmonization efforts, i.e., responses from the developers regarding harmonization.
- Briefly summarize next steps according to protocol
- 0067 : Chronic Stable Coronary Artery Disease: Antiplatelet Therapy
- 0076 : Optimal Vascular Care
- 0142 : Aspirin prescribed at discharge for AMI
- 2452 :Percutaneous Coronary Intervention (PCI): Post-Procedural Optimal Medical Therapy
- 0964 :Therapy with Aspirin, P2Y12 Inhibitor, and Statin at Discharge Following PCI in Eligible Patients
- 0465: Perioperative Anti-platelet Therapy for Patients undergoing Carotid Endarterectomy

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0068

Measure Title: Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet

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: 6/30/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Routine use of aspirin or another antiplatelet

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Patient 18 years of age or older >>>> Diagnosed with cardiovascular disease (either through having an event like an AMI, CABG, or PCI or by having a diagnosis consistent with ischemic vascular disease in the outpatient setting) >>>> Health provider discusses benefits of antiplatelet therapy with patient >>>> Health care provider prescribes aspirin or another antiplatelet >>>> Patient uses antiplatelet medication routinely >>>> Patient has significant reduction in risk of having an AMI, stroke, vascular complication, or dying.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Smith SC Jr, Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, Gibbons RJ, Grundy SM, Hiratzka LF, Jones DW, Lloyd-Jones DM, Minissian M, Mosca L, Peterson ED, Sacco RL, Spertus J, Stein JH, Taubert KA. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation. 2011;124: 00-00.

- http://circ.ahajournals.org/content/124/22/2458

Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, Fang MC, Fisher M, Furie KL, Heck DV, Johnston SC, Kasner SE, Kittner SJ, Mitchell PH, Rich MW, Richardson D, Schwamm LH, Wilson JA; on behalf of the American Heart Association Stroke Council, Council on Cardiovascular and Stroke Nursing, Council on Clinical Cardiology, and Council on Peripheral Vascular Disease. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2014;45:2160–2236.

- http://stroke.ahajournals.org/content/early/2014/04/30/STR.0000000000000024

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

AHA/ACCF Secondary Prevention and Risk Reduction Therapy for Patients with Coronary and Other Atherosclerotic Vascular Disease: 2011 Update

Page Number	Recommendation	Verbatim Quote
3	1	Aspirin 75–162 mg daily is recommended in all patients with coronary artery disease unless contraindicated. (Level of Evidence: A) • Clopidogrel 75 mg daily is recommended as an alternative for patients who are intolerant of or allergic to aspirin. (Level of Evidence: B) Class I
3	2	A P2Y12 receptor antagonist in combination with aspirin is indicated in patients after ACS or PCI with stent placement. (Level of Evidence: A) • For patients receiving a bare-metal stent or drug-eluting stent during PCI for ACS, clopidogrel 75 mg daily, prasugrel 10 mg daily, or ticagrelor 90 mg twice daily should be given for at least 12 months. (Level of Evidence: A) Class I
4	3	For patients undergoing coronary artery bypass grafting, aspirin should be started within 6 hours after surgery to reduce saphenous vein graft closure. Dosing regimens ranging from 100 to 325 mg daily for 1 year appear to be efficacious. (Level of Evidence: A) Class I

Page Number	Recommendation	Verbatim Quote
4	4	In patients with extracranial carotid or vertebral atherosclerosis who have had ischemic stroke or TIA, treatment with aspirin alone (75–325 mg daily), clopidogrel alone (75 mg daily), or the combination of aspirin plus extended-release dipyridamole (25 mg and 200 mg twice daily, respectively) should be started and continued. (Level of Evidence: B) Class I
4	5	For patients with symptomatic atherosclerotic peripheral artery disease of the lower extremity, antiplatelet therapy with aspirin (75–325 mg daily) or clopidogrel (75 mg daily) should be started and continued. (Level of Evidence: A) Class I
4	6	Antiplatelet therapy is recommended in preference to anticoagulant therapy with warfarin or other vitamin K antagonists to treat patients with atherosclerosis. (Level of Evidence: A) Class I

Guidelines for the Prevention of Stroke in Patients with Stroke and Transient Ischemic Attack: 2014

Page Number	Recommendation	Verbatim Quote
2198	1	For patients with noncardioembolic ischemic stroke or TIA, the use of antiplatelet agents rather than oral anticoagulation is recommended to reduce the risk of recurrent stroke and other cardiovascular events (Class I; Level of Evidence A).
2198	2	Aspirin (50–325 mg/d) monotherapy (Class I; Level of Evidence A) or the combination of aspirin 25 mg and extended-release dipyridamole 200 mg twice daily (Class I; Level of Evidence B) is indicated as initial therapy after TIA or ischemic stroke for prevention of future stroke. (Revised recommendation)
2198	3	Clopidogrel (75 mg) monotherapy is a reasonable option for secondary prevention of stroke in place of aspirin or combination aspirin/dipyridamole (Class IIa; Level of Evidence B). This recommendation also applies to patients who are allergic to aspirin.
2199	4	For patients with noncardioembolic ischemic stroke or TIA, the use of antiplatelet agents rather than oral anticoagulation is recommended to reduce the risk of recurrent stroke and other cardiovascular events (Class I; Level of Evidence A).

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

AHA/ACCF Secondary Prevention and Risk Reduction Therapy for Patients with Coronary and Other Atherosclerotic Vascular Disease: 2011 Update

Recommendation	Grade	Definition
1-6	Class I	Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

Guidelines for the Prevention of Stroke in Patients with Stroke and Transient Ischemic Attack: 2014

Recommendation	Grade	Definition
1, 2, 4	Class I	Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered
3	Class IIa	Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Grade	Definition
Class IIb	Benefit ≥ Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED
Class III – No Benefit	No Benefit Procedure/Test -Not helpful Treatment – No proven benefit
Class III - Harm	Harm Procedure/Test – Excess cost without benefit or harmful Treatment – Harmful to patients

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Both guidelines cited in 1a.4.1 use the same methodology for grading recommendations.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Both clinical guidelines cited in 1a.4.1 cover a much broader topic area (i.e., secondary prevention) than just antiplatelet therapy and do not discuss in detail the evidence review process used for each antiplatelet therapy recommendation. They do, however, provide a grade of the evidence for each of the recommendations and cite multiple studies, other guidelines, and systematic reviews supporting those recommendations. Therefore, to answer the questions in 1a.7, we are using the evidence grades provided in the guidelines and referencing one seminal systematic review cited in the guidelines that summarizes the body of evidence supporting the recommendations.

- Antithrombotic Trialists' Collaboration Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. BMJ. 2002;324:71–86.

  http://www.bmj.com/content/324/7329/71.long

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Both guidelines cited in 1a.4.1 use the same methodology for grading evidence.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Antithrombotic Trialists' Collaboration Meta-Analysis

This meta-analysis of randomized trials sought to determine the effects of antiplatelet therapy among patients at high risk of occlusive vascular events and studied non-fatal myocardial infarctions, non-fatal stroke (subdivided into intracranial hemorrhages and strokes of ischemic or unknown etiology), and vascular death as main outcomes.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

AHA/ACCF Secondary Prevention and Risk Reduction Therapy for Patients with Coronary and Other Atherosclerotic Vascular Disease: 2011 Update

Recommendation	Grade	Definition
1-3, 5, 6	Level A	Multiple populations evaluated* Data derived from multiple randomized clinical trials or meta-analyses
4, sub-bullet of recommendation 1	Level B	Limited populations evaluated* Data derived from a single randomized trial or nonrandomized studies

* Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use.

Guidelines for the Prevention of Stroke in Patients with Stroke and Transient Ischemic Attack: 2014

Recommendation	Grade	Definition
1, 4	Level A	Multiple populations evaluated* Data derived from multiple randomized clinical trials or

		meta-analyses
2, 3	Level B	Limited populations evaluated* Data derived from a single randomized trial or nonrandomized studies

* Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Grade	Definition
Level C	Very limited populations evaluated* Only consensus opinion of experts, case studies, or standar of care

*Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

The antiplatelet recommendations from the 2011 ACC/AHA Secondary Prevention Guidelines and the 2014 AHA/ASA Stroke Prevention Guidelines cite studies, systematic reviews, and other guidelines as recent as 2013 and dating back to 1978. It should be noted that the search parameters described by some of the systematic reviews did not include clear date ranges (e.g., "all studies published by 1997") so the body of evidence potentially goes back further.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The body of evidence supporting antiplatelet therapy as secondary preventive treatment in patients with established cardiovascular disease is comprised of many (>100) studies, meta-analyses, and practice guidelines.

The Antithrombotic Trialists' Collaboration (ATT Collaboration) ultimately included 287 studies involving 135,000 patients in comparisons of antiplatelet therapy versus control and 77,000 in comparisons of different antiplatelet regimens to make their conclusions. The ATT Collaboration Meta-Analysis reviewed randomized trials of an antiplatelet regimen versus control or of one antiplatelet regimen versus another in high risk patients (with acute or previous vascular disease or some other predisposing condition) from which results were available before September 1997. Trials had to use a method of randomization that precluded prior knowledge of the next treatment to be allocated and comparisons had to be un-confounded—that is, have study groups that differed only in terms of antiplatelet regimen.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

The ATT Collaboration found sufficient quality of evidence across the studies included in their analysis to support the use of antiplatelet therapy as secondary preventive treatment in patient with established cardiovascular disease. Antiplatelet therapy clearly reduced the incidence of non-fatal myocardial infarctions, non-fatal stroke, or vascular death.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Aspirin or other oral antiplatelet drugs are protective in most types of patients at increased risk of occlusive vascular events, including those with an acute myocardial infarction or ischemic stroke, unstable or stable angina, previous myocardial infarction, stroke or cerebral ischemia, peripheral arterial disease, or atrial fibrillation. Overall, among these high risk patients, allocation to antiplatelet therapy reduced the combined outcome of any serious vascular event by about one quarter; non-fatal myocardial infarction was reduced by one third, non-fatal stroke by one quarter, and vascular mortality by one sixth (with no apparent adverse effect on other deaths). Absolute reductions in the risk of having a serious vascular event were 36 per 1000 treated for two years among patients with previous myocardial infarction; 38 per 1000 patients treated for one month among patients with acute myocardial infarction; 36 per 1000 treated for two years among those with previous stroke or transient ischemic attack; 9 per 1000 treated for three weeks among those with acute stroke; and 22 per 1000 treated for two years among other high risk patients (with separately significant results for those with stable angina ($P = 0.0005$), peripheral arterial disease ($P = 0.004$), and atrial fibrillation ($P = 0.01$)). In each of these high risk categories, the absolute benefits substantially outweighed the absolute risks of major extracranial bleeding.

Aspirin was the most widely studied antiplatelet drug, with doses of 75-150 mg daily at least as effective as higher daily doses. The effects of doses lower than 75 mg daily were less certain. Clopidogrel reduced serious vascular events by 10% compared with aspirin, which was similar to the 12% reduction observed with its analogue ticlopidine. Addition of dipyridamole to aspirin produced no significant further reduction in vascular events compared with aspirin alone. Among patients at high risk of immediate coronary occlusion, short term addition of an intravenous glycoprotein IIb/IIIa antagonist to aspirin prevented a further 20 vascular events per 1000 ($P < 0.0001$) but caused 23 major (but rarely fatal) extracranial bleeds per 1000.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Because all antiplatelet drugs interfere with normal blood clotting, the most common side effect or risk associated with using them is bleeding. Bleeding risks range from the very minor, such as nose bleeds, to major, life-threatening events such as bleeding into the brain.

In addition to the main outcomes of non-fatal myocardial infarctions, non-fatal stroke (subdivided into intracranial hemorrhages and strokes of ischemic or unknown etiology), and vascular death, the ATT Collaboration Meta-Analysis also studied extracranial bleeding as a potential harm. The benefits of antiplatelet therapy were found to far exceed that harm (as well as the others) in patients at an increased risk of occlusive vascular events.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Since the publication of the 2002 ATT Collaboration Meta-Analysis, there have been many (>100) studies and systematic reviews focused on the use of antiplatelet therapy as secondary prevention, none of which conflict with the ATT Collaboration findings.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form Final_Evidence_Form_0068_IVD_Aspirin.docx
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., the benefits or improvements in quality envisioned by use of this measure)* This measure aims to improve the quality of care for patients with established cardiovascular disease by assessing whether or not they received aspirin or another antiplatelet medication during the measurement year. Antiplatelet medications, such as aspirin and clopidogrel, are drugs that inhibit platelets from clumping together and forming clots. Their use in the secondary prevention of cardiovascular events is well-established. In patients who are at high risk because they already have atherosclerotic cardiovascular disease, antiplatelet therapy reduces the yearly risk of serious vascular events (MI, stroke, death) by about twenty-five percent.

Antiplatelet agents also have a beneficial effect in reducing all-cause mortality and fatal cardiovascular events in patients with peripheral arterial disease.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

NCQA HEART/STROKE RECOGNITION PROGRAM

The data references are extracted from NCQA's Heart/Stroke Recognition Program reflecting the most recent years of reporting for this measure. Below is a description of the total number of clinicians (N) submitting data for this measure. The figure N represents the total number of individual clinicians reporting data. Performance data is summarized by the mean, standard deviation, performance percentiles (10th, 25th, 50th, 75th and 90th percentile) and the interquartile range.

Performance Data from NCQA's Heart/Stroke Recognition Program – Individual Clinician Level

YEAR	N	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2012	755	65%	24%	38%	44%	59%	89%	97%	45%
2013	659	84%	20%	50%	83%	91%	97%	100%	14%

CMS PHYSICIAN QUALITY REPORTING SYSTEM (PQRS)

The following PQRS data available from CMS includes mean performance for program years 2010-2013. Note that the figure N here represents the total number of eligible professionals reporting data each year for the IVD: Use of Aspirin or Another Antiplatelet measure.

YEAR	N	MEAN
2010	4491	75%
2011	8167	80%
2012	18291	55%
2013	17935	64%

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

N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

See response in question 1b.5

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

Heart disease is the leading cause of death for people of most ethnicities in the United States, including African Americans, Hispanics, and whites. For American Indians or Alaska Natives and Asians or Pacific Islanders, heart disease is the second leading cause of death (CDC, 2015). Non-Hispanic black adults are at least 50% more likely to die of heart disease or stroke prematurely (i.e., before age 75 years) than their non-Hispanic white counterparts (CDC, 2013). Black women and men are more likely to die before age 75 as a result of coronary heart disease (CHD) than white women and men (rates of death are 37.9%, 61.5%, 19.4%, and 41.5%, respectively) (CDC, 2011). Racial and age-related disparities also exist in rates of recurrent MI or fatal CHD within 5 years of a first MI. Of those who have a first MI, the percentage with a recurrent event is as follows: at 45 to 64 years of age, 14% of white men, 18% of white women, 22% of black men, and 28% of black women; at =65 years of age, 21% of white men and women, 33% of black men, and 26% of black women (Mozaffarian et al., 2015).

According to data from the 2013 Behavioral Risk Factor Surveillance System (BRFSS), 2.7% of men and 2.7% of women =18 years of age had a history of stroke; 2.5% of non-Hispanic whites, 4.0% of non-Hispanic blacks, 1.3% of Asian/Pacific Islanders, 2.3% of Hispanics (of any race), 4.6% of American Indian/Alaska Natives, and 4.6% of other races or multiracial people had a history of stroke. Over the time period 2006 to 2010, data from BRFSS show that the overall self-reported stroke prevalence did not change. Older adults, blacks, people with lower levels of education, and people living in the southeastern United States had higher stroke

prevalence (BRFSS, 2013).

With regard to peripheral arterial disease (PAD), the highest prevalence has been observed among elderly people, non-Hispanic blacks, and women (Mozaffarian et al., 2015). The 2011 overall any-mention age-adjusted death rate for PAD was 18.1 per 100,000. Any-mention death rates in males were 21.6 for whites, 24.7 for blacks, 8.8 for Asians or Pacific Islanders, and 16.7 for American Indians or Alaska Natives. In females, rates were 15.7 for whites, 18.3 for blacks, 6.9 for Asians or Pacific Islanders, and 13.0 for American Indians or Alaska Natives (NCHS, 2011).

There are limited studies examining the use and adherence of aspirin because pharmacy claims do not include aspirins and other "over-the-counter" drugs. However, a 2010 study using in-home interviews of 3,005 community-residing individuals, ages 57–85 years, found that blacks are less likely than whites to use aspirin (29% vs. 44%, p = 0.008). After controlling for age, gender, comorbidity, socioeconomic status, and access to care factors, racial/ethnic disparities persisted. In particular, black participants at the highest risk of having a cardiovascular event were less likely than their white counterparts to use aspirin (odd ratio 0.61, CI 0.37–0.98) (Qato et al., 2010).

Using data from the American College of Cardiology's PINNACLE Registry, researchers examined the association between socioeconomic status (SES) and guidelines-based treatment with anti-platelet and statin medications and found that PAD patients with lower SES are less likely to receive antiplatelet and statin therapy. The practice site at which patients received care largely explained the observed SES differences in treatment with guideline-recommended secondary prevention medications. Future efforts to reduce treatment disparities in these vulnerable populations should target systems improvement at practices serving high proportions of patients with low SES. (Subherwal et al., 2013).

Behavioral Risk Factor Surveillance System: annual survey data, 2013. Centers for Disease Control and Prevention Web site. http://www.cdc.gov/brfss/annual_data/annual_2013.html. Accessed June 10, 2015.

Center for Disease Control and Prevention (CDC). 2015. Heart Disease Facts. Last modified February 19, 2015. http://www.cdc.gov/heartdisease/facts.htm

Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services. 2013. "CDC Health Disparities and Inequalities Report-United States, 2013." Morbidity and Mortality Weekly Report (MMWR) 62(03); 1-2. http://www.cdc.gov/mmwr/pdf/other/su6203.pdf

Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services. 2011. "Fact Sheet: Health Disparities in Coronary Heart Disease and Stroke." http://www.cdc.gov/minorityhealth/CHDIR/2011/FactSheets/CHDStroke.pdf

Mozaffarian, D., Benjamin E.J., Go A.S., et al. 2015. "Heart disease and stroke statistics—2015 update: a report from the American Heart Association." Circulation. 131:e29-e322. doi: 10.1161/CIR.0000000000000152

National Center for Health Statistics (NCHS). Public use data sets for final US 2011 mortality tabulated by the National Heart, Lung, and Blood Institute. Mortality multiple cause-of-death public use record. http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm#Mortality_Multiple. Accessed July 3, 2014.

Qato DM, Lindau ST, Conti RM, Schumm LP, Alexander GC.: Racial and ethnic disparities in cardiovascular medication use among older adults in the United States. Pharmacoepidemiol Drug Saf 19: 834–842, 2010.

Subherwal S, Patel MR, Tang F, Smolderen KG, Jones WS, Tsai TT, Ting HH, Bhatt DL, Spertus JA, Chan PS. Socioeconomic disparities in the use of cardioprotective medications among patients with peripheral artery disease: an analysis of the American College of Cardiology´s NCDR PINNACLE Registry. J Am Coll Cardiol. 2013; 62:51-57.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Cardiovascular disease, including stroke, is the leading cause of death in the United States. More than 85 million American adults have one or more types of cardiovascular disease. Specifically, more than 15 million adults (20 years and older) have coronary heart disease (CHD), over 8 million adults have angina, more than 7 million adults have had a myocardial infarction (MI), over 6 million adults have had a stroke, and nearly 7 million adults 40 years of age and older have peripheral artery disease (Mozaffarian et al., 2015). It is estimated that by 2030 more than 43 percent of Americans will have a form of cardiovascular disease (Heidenreich et al., 2011).

In 2011, the total cost of cardiovascular disease and stroke in the United States was estimated to be $320 billion. This total includes direct costs such as the cost of physicians and other health professionals, hospital services, prescribed medications and home health care, as well as indirect costs due to loss of productivity from premature mortality (Mozaffarian et al., 2015). By 2030, direct medical costs for cardiovascular disease are projected to increase to nearly $918 billion (Heidenreich, 2011).

Antiplatelet medications, such as aspirin and clopidogrel, are drugs that inhibit platelets from clumping together and forming clots. Their use in the secondary prevention of cardiovascular events is well established. In patients who are at high risk because they already have occlusive cardiovascular disease, long-term antiplatelet therapy reduces the yearly risk of serious vascular events (MI, stroke, death) by about twenty-five percent (Antiplatelet Trialists´ Collaboration, 1994; 2002; 2009). A more recent systematic review of the literature confirmed the benefits of antiplatelet therapy in reducing death from cardiovascular causes, MI, or stroke (Cheng, 2013). Antiplatelet agents also have a beneficial effect in reducing all-cause mortality and fatal cardiovascular events in patients with peripheral arterial disease (Wong et al., 2011).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Antiplatelet Trialists´ Collaboration. Collaborative overview of randomised trials of antiplatelet therapy. I. Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. BMJ. 1994;308:81–106.

Antithrombotic Trialists´ Collaboration Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. BMJ. 2002;324:71–86.

Antithrombotic Trialists' (ATT) Collaboration; Baigent C, Blackwell L, Collins R, Emberson J, Godwin J, Peto R, Buring J, Hennekens C, Kearney P, Meade T, Patrono C, Roncaglioni MC, Zanchetti A. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta analysis of individual participant data from randomized trials. Lancet. 2009;373:1849 –1860.

Cheng JW. Updates in antiplatelet agents used in cardiovascular diseases. J Cardiovas Pharmacol Ther. 2013;18(6):514–524.

Heidenreich, P.A., J.G. Trogdon, O.A. Khavjou, et al. 2011. "Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association." Circulation.123:933-944.

Mozaffarian, D., E.J. Benjamin, A.S. Go, et al. 2015. "Heart disease and stroke statistics—2015 update: a report from the American Heart Association." Circulation. 131:e29-e322. doi: 10.1161/CIR.0000000000000152

Wong PF, Chong LY, Mikhailidis DP, Robless P, Stansby G. Antiplatelet agents for intermittent claudication. Cochrane Database of Systematic Reviews 2011, Issue 11. Art. No.: CD001272. DOI: 10.1002/14651858.CD001272.pub2.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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 : Acute Myocardial Infarction, Cardiovascular : Ischemic Heart Disease, Coronary Artery Disease, Cardiovascular : Percutaneous Coronary Intervention (PCI), Prevention

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

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.)*
http://www.ncqa.org/tabid/59/Default.aspx

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** 0068_IVD_Value_Sets_Final.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Since this measure last underwent endorsement maintenance, we revised the measure in the following ways:

-Changed “antithrombotic” to “antiplatelet” throughout the measure specification to more closely align with measure intent, which is to assess the use of aspirin or other antiplatelet medications in patients with ischemic vascular disease.

-Added exclusion for patients on anticoagulant therapy based on feedback from measurement advisory panels.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who had documentation of routine use of aspirin or another antiplatelet during the measurement year.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator: The measurement year (12 month period).
Denominator: The measurement year and the year prior to the measurement year (24 month period).

S.6. **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, 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.
ADMINISTRATIVE

Patients who had documentation of routine use of aspirin or another antiplatelet during the measurement year.

Refer to Table IVD-E to identify medications for oral anti-platelet therapy.

ORAL ANTI-PLATELET THERAPIES (TABLE IVD-E)
PRESCRIPTIONS
- Aspirin
- Clopidogrel
- Aspirin-dipyridamole
- Prasugrel
- Ticagrelor
- Ticlopidine

MEDICAL RECORD
Patients who had documentation of routine use of aspirin or another antiplatelet during the measurement year.

At a minimum, documentation in the medical record must include a note indicating the date when aspirin or another antiplatelet was prescribed or documentation of prescription from another treating physician.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients 18 years or older by the end of the measurement year discharged from an inpatient setting with an AMI, CABG, or PCI during the 12 months prior to the measurement year or who had a diagnosis of IVD during both the measurement year and the year prior to the measurement year.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

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

Patients are identified for the eligible population in two ways: by event or by diagnosis. The organization must use both methods to identify the eligible population, but a patient only needs to be identified by one method to be included in the measure.

Event. Any of the following during the year prior to the measurement year meet criteria:
- MI. Discharged from an inpatient setting with an MI (MI Value Set)*. Use both facility and professional claims to identify MI.
-CABG. Discharged from an inpatient setting with a CABG (CABG Value Set)*. Use both facility and professional claims to identify CABG.
-PCI. Patients who had a PCI (PCI Value Set)* in any setting.

Diagnosis. Patients who meet at least one of the following criteria during both the measurement year and the year prior to the measurement year. Criteria need not be the same across both years.
-At least one outpatient visit (Outpatient Value Set)* with an IVD diagnosis (IVD Value Set)*, or
-At least one acute inpatient encounter (Acute Inpatient Value Set)* with an IVD diagnosis (IVD Value Set)*.

*Due to the extensive volume of codes associated with identifying the denominator for this measure, we are attaching a separate file with code value sets. See code value sets located in question S.2b.

MEDICAL RECORD
Documentation of IVD in the medical record includes:
- IVD
- Ischemic heart disease
- Angina
- Coronary atherosclerosis
- Coronary artery occlusion

- Cardiovascular disease
- Occlusion or stenosis of precerebral arteries (including basilar, carotid and vertebral arteries)
- Atherosclerosis of renal artery
- Atherosclerosis of native arteries of the extremities
- Chronic total occlusion of artery of the extremities
- Arterial embolism and thrombosis
- Atheroembolism.

Note: Use paper logs, patient registries or electronic medical records (EMRs) to identify the denominator, then use the medical record to confirm patient eligibility.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients who had documentation of use of anticoagulant medications during the measurement year.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Patients who had documentation of use of anticoagulant medications during the measurement year.

ANTICOAGULANT MEDICATIONS
- Apixaban
- Argatroban
- Bivalirudin
- Dabigatran
- Dalteparin
- Desirudin
- Edoxaban
- Enoxaparin
- Fondaparinux
- Heparin
- Lepirudin
- Rivaroxaban
- Tinzaparin
- Warfarin

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Step 1: Determine the denominator
Patients 18 years of age or older by the end of the measurement year AND who were discharged from an inpatient setting for an AMI, CABG or PCI during the 12 months prior to the measurement year or who had a diagnosis of IVD during both the measurement year and the year prior to the measurement year.

Step 2: Exclude patients who meet the exclusion criteria
Patients on anticoagulant therapy.

Step 3: Determine the numerator
Patients who had documentation of routine use of aspirin or another antiplatelet during the measurement year.

Step 4: Calculate the rate by dividing the numerator (Step 3) by the denominator (after exclusions) (Step 2).

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
N/A

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Final_Testing_Form_0068_IVD_Aspirin-635712772001794970.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0068
Measure Title: Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
Date of Submission: 6/30/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

1.3. What are the dates of the data used in testing?
Testing of reliability of the performance measure score with beta binomial was performed with data from measurement year 2013.

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:	**Measure Tested at Level of:**

(*must be consistent with levels entered in item S.26*)	
☒ 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*)

ENTITIES INCLUDED FOR MEASURE SCORE RELIABILTY TESTING
Measure score reliability was calculated from NCQA's Heart/Stroke Recognition Program (HSRP) that included 659 clinicians who voluntarily participate in the program The HSRP recognizes clinicians and group practices for the delivery of quality ambulatory care to persons with cardiovascular disease.

SYSTEMATIC ASSESSMENT OF FACE VALIDITY
This measure was tested for face validity with two expert panels. See additional information: Ad.1. Workgroup/Expert Panel in Measure Development for names and affiliation of expert panels:

1. Cardiovascular Measurement Advisory Panel includes eight physicians and one nurse with expertise in cardiovascular health and quality measurement.
2. NCQA's Clinical Programs Committee (CPC) oversees the evolution of NCQA's recognition programs and related measures including the Diabetes Recognition Program, the Heart/Stroke Recognition Program, the Patient Centered Medical Home and Patient-Centered Specialty Practice Recognition Program, among others. The CPC includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 15 members. The CPC is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of clinical recognition programs. CPC members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

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

HEART/STROKE RECOGNITION PROGRAM
Clinicians submitting data to the Heart Stroke Recognition Program establish a start date for abstraction and working consecutively backward from that date, identify 35 patients who had office visits and meet three eligibility criteria:

1. 18 years of age or older
2. Diagnosis of ischemic vascular disease (e.g., coronary artery disease, stroke, peripheral arterial disease, etc.) for at least 12 months
3. Under the care of the applicant clinician or group practice for IVD for at least 12 months. This is defined by documentation of a face-to-face visit for IVD care between the clinician and the patient that pre-dates the most recent IVD visit by at least 12 months.

Clinicians report clinical data relevant to the HSRP measures from the 35 patient records that met the eligibility criteria to NCQA for review. In 2013, a total of 659 clinicians submitted data on approximately 23065 patients.

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

Reliability of this measure was tested using date from 2013. Validity was tested through a systematic assessment of face validity. We have described the composition of the expert panels that assessed face validity in the data sample questions above.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
N/A

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

METHOD FOR PERFORMANCE MEASURE SCORE RELIABILITY TESTING (BETA-BINOMIAL):
In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009). The following is quoted from the tutorial which focused on provider-level assessment, "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient."

The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance across accountable entities.

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

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

Below, we present available statistics from data submitted by 659 clinicians in the Heart/Stroke Recognition Program for the IVD: Use of Aspirin or Another Antiplatelet measure.

Overall Reliability	10th Percentile-90th Percentile
0.88	0.82-1.00

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.e., *what do the results mean and what are the norms for the test conducted?*)
PERFORMANCE MEASURE SCORE (BETA BINOMIAL)
Reliability scores can vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (signal). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities.

Testing suggests that this measure has strong overall reliability at 0.88. The 10th-90th percentile distribution of physician level reliability on the rates in this measure show the vast majority of physicians met or exceeded the minimally accepted threshold of 0.7, and the majority of them exceeded 0.8. Strong reliability is demonstrated since majority of variances is due to signal and not to noise.

2b2. VALIDITY TESTING
2b2.1. 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*)

2b2.2. 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)*

SYSTEMATIC ASSESSMENT OF FACE VALIDITY
Method of Assessing Face Validity: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. The work-up is vetted by NCQA's Measurement Advisory Panels (MAPs) and the Clinical Programs Committee (CPC), as well as other panels as necessary.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with practices or health plans to conduct field-tests that assess the feasibility and validity of potential measures. NCQA uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA about new measures or about changes to existing measures. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations. New measures and changes to existing measures approved by the CPC and NCQAs Board of Directors will be included in the next recognition program year.

STEP 4: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, NCQA prioritizes measures for re-evaluation and selected measures are researched for changes in clinical guidelines or in the health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be recommended for retirement. The CPC reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's measure set.

Expert Participation

The IVD: Use of Aspirin or Another Antiplatelet was tested for face validity with two expert panels. Updated guidelines from the American Heart Association, the American College of Cardiology, and the American Stroke Association were also strong authoritative sources in re-evaluating the IVD: Use of Aspirin or Another Antiplatelet measure.

We list each panel here. See additional information: Ad.1. Workgroup/Expert Panel in Measure Development for names and affiliation of expert panels:

1. Cardiovascular Measurement Advisory Panel includes eight physicians and one nurse with expertise in cardiovascular health and quality measurement.
2. NCQA's Clinical Programs Committee (CPC) oversees the evolution of NCQA's recognition programs and related measures including the Diabetes Recognition Program, the Heart/Stroke Recognition Program, the Patient Centered Medical Home and Patient-Centered Specialty Practice Recognition Program, among others. The CPC includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 17 members. The CPC is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of clinical recognition programs. CPC members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

N/A

2b2.4. 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?*)

SYSTEMATIC ASSESSMENT OF FACE VALIDITY
The results indicate that the multiple NCQA panels concluded with good agreement that the measure, as specified accurately differentiates quality across clinicians and group practices. Our interpretation of these results is that IVD: Use of Aspirin or Another Antiplatelet measure is a valid and necessary quality metric.

2b3. EXCLUSIONS ANALYSIS
NA ☒ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Prior to 2015, the IVD: Use of Aspirin or Another Antiplatelet measure did not have any exclusions.

2b3.2. 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*)
N/A
2b3.3. 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*)
N/A

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected entities (e.g., physicians, clinicians, plans) at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each entity. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than .05, then the two entities' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected physicians, one in the 25th percentile and another in the 75th percentile of performance. We used these two physicians as examples of measured entities. However, the method can be used for comparison of any two measured entities.

2b5.2. 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*)

Variation in Performance Across Physicians in the Heart/Stroke Recognition Program

Year	N	Mean	SD	10th	25th	50th	75th	90th	IQR	p-value
2013	659	84%	20%	50%	83%	91%	97%	100%	14%	<0.05

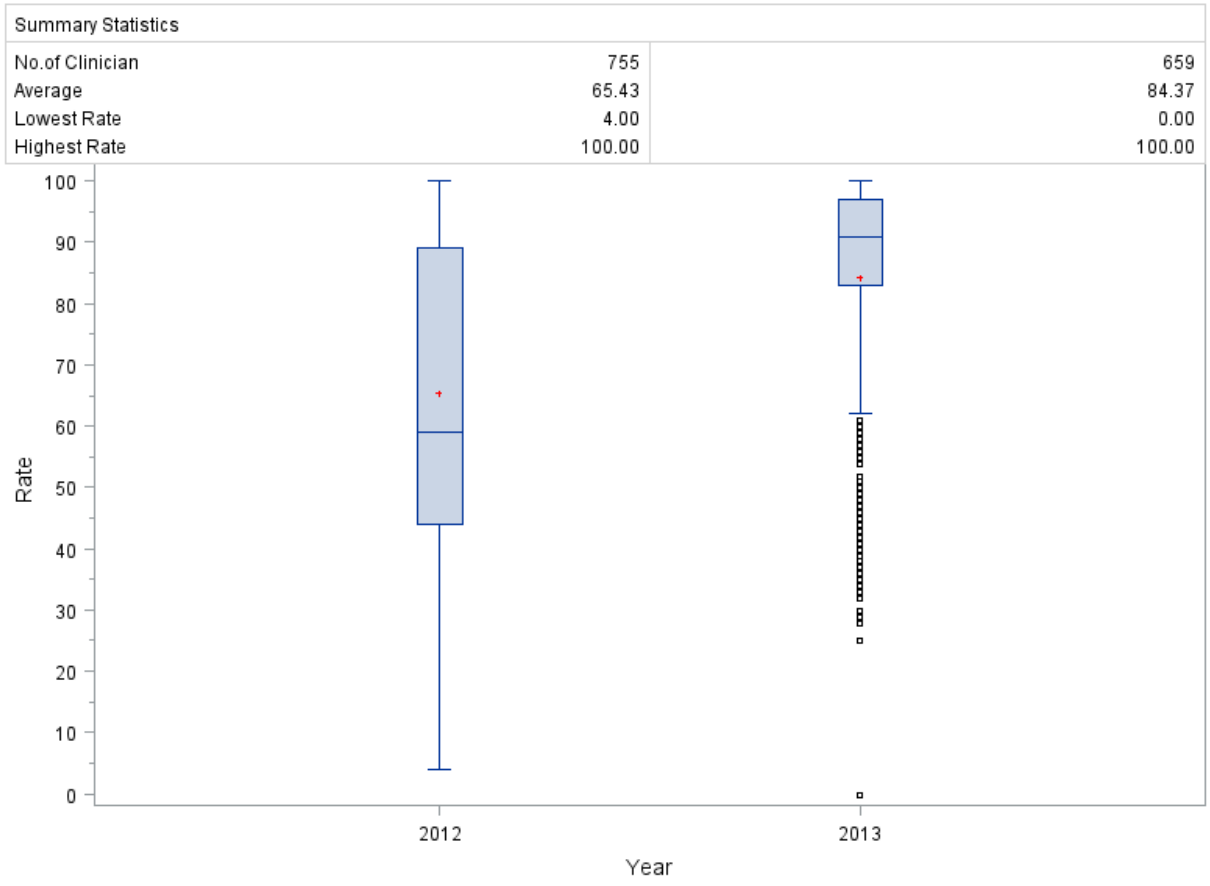
N = total number of physicians reporting data for recognition

IQR: Interquartile range

p-value: p-value of independent samples t-test comparing physicians at the 25th percentile to physicians at the 75th percentile

A box plot of the Heart/Stroke Recognition Program (HSRP) data from 2012 and 2013 is included below for your reference. The HSRP includes physicians who voluntarily submit data on a sample of 35 ischemic vascular disease patients. Please note the title of the measure changed to "Use of Aspirin or Another Antiplatelet" in 2015.

Ischemic Vascular Disease (IVD): Use of Aspirin or another Antithrombotic from 2012-2013 Clinician level



2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across

measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The results indicate there is a 14% gap in performance between physicians in the 25th and 75th percentiles and the difference is statistically significant. There is also a 50% gap in performance between the 10th and 90th percentiles. Overall, results suggest there are meaningful differences in performance and there is an opportunity for improvement. We would expect performance of clinicians outside of the HSRP program to show wider variation.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
N/A

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)
This measure is collected with a complete sample, there is no missing data on this measure.

2b7.2. 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*)
This measure is collected with a complete sample, there is no missing data on this measure.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how

the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)
This measure is collected with a complete sample, there is no missing data on this measure.

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)*
Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**
To allow for widespread reporting across health care practices, this measure is collected through multiple data sources (administrative data, electronic clinical data, and paper records). We anticipate as electronic health records become more widespread, the reliance on paper record review will decrease.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
We have found that physicians participating in NCQA's Heart/Stroke Recognition Program have no difficulty collecting, interpreting and reporting the data for this measure. We receive constant feedback from the field during our reviews which helps us strengthen the program. We have received positive feedback on the use of this measure in the CMS PQRS program and few questions have been raised by participating clinicians to the CMS vendor.

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

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

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.

Planned	Current Use (for current use provide URL)
	Public Reporting CMS Physician Quality Reporting System (PQRS) http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/ Payment Program CMS EHR Incentive Program (Meaningful Use) http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html Regulatory and Accreditation Programs CMS ACO Shared Savings Program http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Quality_Measures_Standards.html Professional Certification or Recognition Program NCQA Heart/Stroke Recognition Program http://www.ncqa.org/tabid/140/Default.aspx

4a.1. For each CURRENT use, checked above, provide:

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

CMS PHYSICIAN QUALITY REPORTING SYSTEM: This measure is used in the Physician Quality Reporting System (PQRS) which is a reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs). PQRS is a voluntary individual reporting program that provides an incentive payment to identified eligible professionals who satisfactorily report data on quality measures for covered Physician Fee Schedule (PFS) services furnished to Medicare Part B beneficiaries (including Railroad Retirement Board and Medicare Secondary Payer). Medicare Part C–Medicare Advantage beneficiaries are not included in claims-based reporting of individual measures or measures groups.
CMS EHR INCENTIVE PROGRAM (MEANINGFUL USE): The Medicare and Medicaid Electronic Health Care Record (EHR) Incentive Programs provide incentive payments to eligible professionals, eligible hospitals, and critical access hospitals (CAHs) as they adopt, implement, upgrade or demonstrate meaningful use of certified EHR technology.
CMS ACO SHARED SAVINGS PROGRAM: The Medicare Shared Savings Program (Shared Savings Program) was established by section

3022 of the Affordable Care Act. The Shared Savings Program is a key component of the Medicare delivery system reform initiatives included in the Affordable Care Act and is a new approach to the delivery of health care. Congress created the Shared Savings Program to facilitate coordination and cooperation among providers to improve the quality of care for Medicare Fee-For-Service (FFS) beneficiaries and reduce unnecessary costs. Eligible providers, hospitals, and suppliers may participate in the Shared Savings Program by creating or participating in an Accountable Care Organization (ACO).
HEART/STROKE RECOGNITION PROGRAM: This measure is used in NCQA's Heart Stroke Recognition Program (HSRP) that assesses clinician performance on key quality measures that are based on national evidence-based guidelines for secondary prevention of cardiovascular disease and stroke. The program includes the following measures: blood pressure control, use of aspirin or another antiplatelet, and smoking status and cessation advice or treatment. The HSRP Recognition provides assurance that clinicians are providing high quality, evidenced–based care for their CVD and stroke patients. Eligible clinicians abstract data from the charts or EHRs of CVD/stroke patients (35 patients for a single applicant) and submit this information to NCQA for review.

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

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

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Refer to data in 1b.2.

HEART/STROKE RECOGNITION PROGRAM
From 2012 to 2013, mean performance increased from 65% to 84%. It should also be noted that performance in each percentile also greatly increased from 2012 to 2013.

PQRS
From 2010 to 2011, mean performance increased from 75% to 80% but then dropped to 55% in 2012. This drop may be due to the sharp increase in the number of providers voluntarily reporting on this measure for the first time in 2012 and not yet progressing up the learning curve. The following year in 2013, mean performance increased to 64%.

4b.2. **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.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We have not identified unintended negative consequences to individuals during the testing and long-standing use of this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0067 : Chronic Stable Coronary Artery Disease: Antiplatelet Therapy
0076 : Optimal Vascular Care
0142 : Aspirin prescribed at discharge for AMI

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
DUE TO THE TEXT LIMIT IN THIS SECTION – WE ARE PROVIDING OUR ANSWER FOR 5a.2 IN SECTION 5b.1.

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
ANSWER FOR SECTION 5a.2

Our current measure, NQF 0068 – Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet, assesses the percentage of patients 18 years of age and older who were discharged from an inpatient setting with an acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary intervention (PCI) during the 12 months prior to the measurement year, AND patients who had a diagnosis of ischemic vascular disease (IVD) during the measurement year and the year prior to the measurement year, who had documentation of the routine use of aspirin or another antiplatelet during the measurement year. NQF 0068 uses administrative claims, electronic clinical data, electronic health record data, and paper medical records from the ambulatory care setting, providing a wide array of options for how data can be collected and reported.

The following is a description of the differences and the impact on interpretability and data collection burden between NQF 0068 and each related measure listed in 5.1a:

NQF 0142 – ASPIRIN PRESCRIBED AT DISCHARGE FOR AMI
This measure assesses the percentage of AMI patients, 18 years and older, who are prescribed aspirin at hospital discharge. The measure population only includes patients who have had an AMI, whereas NQF 0068 includes patients who have had an AMI, CABG or PCI procedure, and patients who have diagnoses consistent with ischemic vascular disease. NQF 0142 focuses only on aspirin prescribed at discharge while NQF 0068 focuses on documentation of the use of any antiplatelet medication during the measurement year. NQF 0142 is a facility-level measure that uses administrative claims and paper medical records from the inpatient setting; NQF 0068 is a physician-level measure that uses administrative claims, electronic clinical data, electronic health record data, and paper medical records from the ambulatory care setting.

There is no impact on interpretability of publically-reported rates or added burden of data collection because the focus of each measure is different, the accountable entity is different and the data for each measure is collected from different data sources by different entities. Additionally, both use value sets of codes to identify patients with AMI that do not conflict.

NQF 0067 – CHRONIC STABLE CORONARY ARTERY DISEASE: ANTIPLATELET THERAPY
This measure assesses the percentage of patients aged 18 years and older with a diagnosis of coronary artery disease (CAD) who were seen by a physician within a 12-month period and who were prescribed aspirin or clopidogrel. The focus of this measure is very similar to NQF 0068 in that it assesses the routine use of antiplatelet therapy in a twelve-month period for patients with CAD. However, NQF 0068 includes more antiplatelet medications than just aspirin or clopidogrel and includes a broader population of patients with cardiovascular disease than just those with CAD.

Although NQF 0067 and NQF 0068 are both physician-level measures that are specified to collect data from administrative claims, electronic clinical data, electronic health record data, and paper medical records from the ambulatory care setting, the impact on interpretability of publically-reported rates or added burden of data collection should be minimal because NQF 0067 is currently only reported through registry data. Additionally, NQF 0067 is focused on only on patients with CAD, while NQF 0068 is focused on a broader population of patients with cardiovascular disease who would benefit from the use of antiplatelet medications.

NQF 0076 – OPTIMAL VASCULAR CARE
This composite measure assesses the percentage of adult patients ages 18 to 75 who have ischemic vascular disease with optimally-managed modifiable risk factors (blood pressure, tobacco-free status, daily aspirin use) at their most recent visit with a physician during the measurement year. While the focus populations for NQF 0076 and NQF 0068 are very similar, NQF 0076 is a composite that includes assessment of blood pressure control and tobacco use status. NQF 0068 assesses the routine use of aspirin or other antiplatelet medications while NQF 0076 focuses only on aspirin use. NQF 0076 does not use administrative claims though it does use electronic clinical data, electronic health record data, and paper medical records from the ambulatory care setting, which is similar to NQF 0068.

Despite the similarities, there should be minimal impact on interpretability of publically-reported rates or added burden of data collection between the two measures since NQF 0076 is a composite of multiple indicators while NQF 0068 is focused only on antiplatelet therapy.

NQF 2452 – PERCUTANEOUS CORONARY INTERVENTION (PCI): POST-PROCEDURAL OPTIMAL MEDICAL THERAPY (NOTE: UNABLE TO SELECT IN 5.a1)
NQF 2452 is a composite measure that assesses the percentage of patients undergoing PCI who receive prescriptions for all medications (aspirin, P2Y12 and statins) for which they are eligible for at discharge. The measure population for NQF 2452 is patients undergoing PCI while NQF 0068 includes patient who have had an AMI, CABG or PCI procedure, and patients who have diagnoses consistent with ischemic vascular disease. NQF 2452 assesses the prescription of aspirin, P2Y12 agents, and statins at discharge; NQF 0068 assesses documentation of use of antiplatelet medications during the measurement year. NQF 2452 is a physician-level measure that uses data from registries while NQF 0068 is a physician-level measure that uses administrative claims, electronic clinical data, electronic health record data, and paper medical records from the ambulatory care setting.

There is no impact on interpretability of publically-reported rates or added burden of data collection because the focus of each measure is different and the data for each measure is collected from different data sources by different entities.

NQF 0964 – THERAPY WITH ASPIRIN, P2Y12 INHIBITOR, AND STATIN AT DISCHARGE FOLLOWING PCI IN ELIGIBLE PATIENTS (NOTE: UNABLE TO SELECT IN 5.a1)
NQF 0964 is a composite measure that assesses the percentage of patients undergoing PCI who receive prescriptions for all medications (aspirin, P2Y12 and statins) for which they are eligible for at discharge. The measure population for NQF 0964 is patients

undergoing PCI while NQF 0068 includes patient who have had an AMI, CABG or PCI procedure, and patients who have diagnoses consistent with ischemic vascular disease. NQF 0964 assesses the prescription of aspirin, P2Y12 agents, and statins at discharge; NQF 0068 assesses documentation of use of antiplatelet medications during the measurement year. NQF 0964 is a facility-level measure that uses data from registries while NQF 0068 is a physician-level measure that uses administrative claims, electronic clinical data, electronic health record data, and paper medical records from the ambulatory care setting.

There is no impact on interpretability of publically-reported rates or added burden of data collection because the focus of each measure is different, the accountable entity is different and the data for each measure is collected from different data sources by different entities.

ANSWER FOR SECTION 5b.1

Our current measure, NQF 0068, has a long history of use and is implemented in four national programs: PQRS, EHR Incentive Program, CMS ACO Shared Savings Program, and the Heart/Stroke Recognition Program.

Appendix

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

Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance
Co.2 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-
Co.3 **Measure Developer if different from Measure Steward:** National Committee for Quality Assurance
Co.4 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
CARDIOVASCULAR MEASUREMENT ADVISORY PANEL
Stephen Persell, MD, MPH (Chair), Northwestern University
Tom Kottke, MD, HealthPartners
Eduardo Ortiz, MD, MPH, Wolters Kluwer Health
David Goff, Jr., MD, PhD, FAHA, FACP, University of Colorado
Kathy Berra, MSN, ANP, FAAN, Stanford University
Michael Pignone, MD, MPH, University of North Carolina at Chapel Hill
Randall S. Stafford, MD, PhD, Stanford University
Tracy Wolff, MD, Agency for Healthcare Research and Quality
Corinne Husten, MD, MPH, U.S. Food and Drug Administration

CLINICAL PROGRAMS COMMITTEE
Randall Curnow, MD, MBA, FACP, FACHE, FACPE (Chair), Mercy Health Physicians
Colleen Conry, MD, University of Colorado
Carol Greenlee, MD, Western Slope Endocrinology
Jennifer Gutzmore, MD, CIGNA
Melissa Hogan, MPH, St. Louis Area Business Health Coalition
Jim Knickman, PhD, New York State Health Foundation
Amy Nguyen, MD, MBA, FAAFP, CAPG

Frank Opelka, MD, FACS, Louisiana State University
Janet Lee Partridge, National Partnership for Women & Families
Keith Reissaus, Nurse-Family Partnership
Marc Rivo, MD, Prestige Health Choice
Nancy Rothman, EdD, RN, Temple University
Julie Schilz, BSN, MBA, Anthem
Xavier Sevilla, MD, FAAP, Catholic Health Initiatives
Lina Walker, PhD, AARP

COMMITTEE ON PERFORMANCE MEASUREMENT
Peter Bach, MD, Memorial Sloan Kettering Cancer Center
Bruce Bagley, MD, TransforMED
Andrew Baskin, MD, Aetna
A. John Blair lll, MD, Taconic IPA, Inc
Patrick Conway, MD, MSC, Center for Medicare & Medicaid Services
Jonathan D. Darer, MD, Geisinger Health System
Helen Darling, National Business Group on Health
Foster Gesten, MD, NYSDOH Office of Managed Care
Marge Ginsburg, Center for Healthcare Decisions
Christine Hunter, MD, (Co-Chair) US Office of Personnel Management
Jeffrey Kelman, MMSc, MD, Centers for Medicare & Medicaid Services
Philip Madvig, MD, The Permanente Medical Group
J. Brent Pawlecki, MD MMM, The Goodyear Tire & Rubber Company
Susan Reinhard, RN, PhD, AARP
Eric C. Schneider, MD, MSc (Co-Chair), RAND Corporation
Marcus Thygeson, MD, MPH Blue Shield of California

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2009

Ad.3 **Month and Year of most recent revision:** 06, 2015

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

Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Ad.8 **Additional Information/Comments:** Publication of each Measure is to be accompanied by the following notice:

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

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA. © 2012 by the National Committee for Quality Assurance.

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 #: 0070
De.2. **Measure Title:** Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)
Co.1.1. **Measure Steward:** AMA-convened Physician Consortium for Performance Improvement
De.3. **Brief Description of Measure:** Percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who also have a prior MI or a current or prior LVEF <40% who were prescribed beta-blocker therapy
1b.1. **Developer Rationale:** For patients with coronary artery disease (CAD), beta-blockers are recommended for 3 years after myocardial infarction or acute coronary syndrome. Beta-blockers, particularly carvedilol, metoprolol succinate, or bisoprolol which have been shown to reduce risk of death, are recommended indefinitely for patients with CAD and LV systolic dysfunction. These agents have proven efficacy in reducing angina onset and improving the ischemic threshold during exercise. In patients who have suffered an MI, beta-blockers significantly reduce deaths and recurrent MIs. (1) Nonadherence to cardioprotective medications is prevalent among outpatients with CAD and can be associated with a broad range of adverse outcomes, including all-cause and cardiovascular mortality, cardiovascular hospitalizations, and the need for revascularization procedures.(2) This measure is intended to promote beta-blocker usage in select patients with CAD.

References:
1. Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, Douglas PS, Foody JM, Gerber TC, Hinderliter AL, King SB III, Kligfield PD, Krumholz HM, Kwong RYK, Lim MJ, Linderbaum JA, Mack MJ, Munger MA, Prager RL, Sabik JF, Shaw LJ, Sikkema JD, Smith CR Jr, Smith SC Jr, Spertus JA, Williams SV. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2012;60:e44-164.
2. Gibbons RJ, Abrams J, Chatterjee K, Daley J, Deedwania PC, Douglas JS, Ferguson TB Jr., Fihn SD, Fraker TD Jr., Gardin JM, O'Rourke RA, Pasternak RC, Williams SV. ACC/AHA 2002 guideline update for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1999 Guidelines for the Management of Patients with Chronic Stable Angina). 2002. Available at: www.acc.org/clinical/guidelines/stable/stable.pdf

S.4. **Numerator Statement:** Patients who were prescribed beta-blocker therapy
S.7. **Denominator Statement:** All patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who also have a prior MI (within the past 3 years) or a current or prior LVEF <40%
S.10. **Denominator Exclusions:** Documentation of medical reason(s) for not prescribing beta-blocker therapy (eg, 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 health care system)

De.1. **Measure Type:**
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry
S.26. **Level of Analysis:** Clinician : Group/Practice, Clinician : Individual
Is this an eMeasure? ☒ **Yes** ☐ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☒ **Yes** ☐ **No**

This measure is submitted as a Registry measure & an eMeasure.

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**

For a MAINTENANCE, what is the Original Endorsement Date: 1/10/09 **Most Recent Endorsement Date:** 1/18/12

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

Public and Member Comments:

- Measure 0071 requires pharmacy data which is not available to clinicians. A clinician-level measure is needed for this process of care. Greater use of low-cost generic medications from discount pharmacies may not be captured in the pharmacy data collection.

Steering Committee: The Committee agreed these issues have merit and re-voted on recommending the measure: Y=8 , N=4 to recommend both 0070 and 0071

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence (including a quantity, quality and consistency assessment) where the evidence focus matches what is being measured, and leads to a desired health outcome.

- This clinician-level process *registry* and *eMeasure* calculates the percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12-month period who also have a prior MI or a current or prior LVEF <40% who were prescribed beta-blocker therapy. Three beta-blockers may be prescribed for patients with LVEF <40%, and all beta-blockers may be prescribed for patients with a prior MI.
- The developer provide the 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS for the guideline for the diagnosis and management of patients with stable ischemic heart disease, using the 2010ACCF/AHA guideline methodology to grade the guideline, (Class I, Level of Evidence: B) for patients with normal LVSD and a previous MI (Class I, Level of Evidence: A) for patients with LVEF < 40%. The guideline data represents patient care from 1996-2009.
- For patients with normal LVSD and a previous MI 3 articles support the recommendation, including 2 systematic reviews & an observational study. For patients with LVEF < 40%, 5 articles are provided including 3 RCTs, 1 meta-analysis of RCTs, and 1 comparative analysis of RCTs. The quality of the articles is not provided. Beta-blockade side effects are discussed as measures harms are noted. An analysis of 75 additional articles since the guideline is also provided.
- A decision logic diagram is provided for beta-blockades to reduce death risk and angina onset, reduce recurrent MIs for patients with previous MIs, and improve ischemic thresholds during exercise.

Questions for the Committee:

- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the*

measure?

- *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- The developer provides average performance data from 2013 PQRS Experience Report which is consistent with 2008-2010 NCDR PINNACLE Registry Data of nearly 90,000 eligible patients:
  - 2010: 71.4%
  - 2011: 82.1%
  - 2012: 69.9%
  - 2013: 74.2% (2013 Small Group Practice Exception Rate: 2.0%)
- PQRS Group Practice Reporting Option (GPRO) Registry reporting volumes are not provided. The developers state that PQRS is a voluntary reporting program. Performance data is not provided from the eMeasure.
- No data on disparities was provided, though the disparities literature finds similar rates between blacks & whites, and genders. Uninsured patients were less likely to receive beta-blockades, though no differences were noted between public and private insurance recipients.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - None.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- The evidence is strong

1b. ***Committee's Comments on Performance Gap:***

- Performance gap remains with 2013 data only hitting 74%

1c. ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- This process measure assess the percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who also have a prior MI or a current or prior LVEF <40% who were prescribed

beta-blocker therapy.

- The developer submits 2 specifications for the measure: Registry & eMeasure. ICD-9 & ICD-10 codes are included, though the ICD-10 conversion methodology is not discussed. In both sets of specifications, the logic is unambiguous. Higher scores equal better quality.
- All eMeasure specifications and values sets meet current NQF technical requirements and are provided on Sharepoint for SC review.
- The measure is intended for use in an office visit, outpatient consultation, nursing facility, long-term care residential facility, home health and provider interaction during the measurement period.
- The developer includes arrhythmia, hypotension, asthma, bradycardia, atrioventricular block and cardiac pacer in Situ as exclusions in the eMeasure specification for both populations (MI within prior 3 years & LVEF <40%), though not in the Registry specification.
- The developer states that exceptions should only be considered when the numerator activity was not performed, that they are not uniformly relevant across measures, and that there must be a clear rationale to permit an exception for a medical, patient, or system reason. In the provided value sets, broadly defined and inappropriate patient, medical and system reason denominator exclusions include but are not limited to: medical reason: 216952002 failure in dosage (event), patient reason: 224187001 variable income (finding) & 266966009 family illness (situation), system reason:266756008 medical care unavailable (situation).
- The eMeasure and Registry Calculation Algorithm are clear, though the denominator details for the Registry specification state a minimum of 2 visits are needed to determine an established patient-provider relationship. Two visits should be included in the description and denominator statement to provide consistent measure calculation. The developer should also clarify if the 2 visits occurred at any time of within the measurement period. Two visits are included in the eMeasure.
- Provider interactions are listed as encounters in the value set spreadsheet and in the eMeasure specifications that include both face-to-face visits and non-face-to-face communications. The developer is encouraged to provide reasoning for inclusion, and clarify if all provider interactions are included in the denominator definition for a patient encounter.
- In addition to critical data element testing, the developer submitted pre-testing from the Measure Authoring Tool within the Bonnie Output that tests eMeasure logic. The measure logic successfully validated through the Bonnie Output.
- The measure is not risk adjusted and SDS variables were not captured for the measure. The developer encourages users to provide collect data and stratify results by race, ethnicity, administrative sex, and payer consistent with CMS' Measures Management System Blueprint and recent national recommendations put forth by the IOM and NQF.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

GPRO Registry Testing

- The developer tested reliability at the performance measure score level, using a beta-binomial model in a signal-to-noise analysis, to differentiate the true differences between measured entities (the signal) from random measurement error (the noise). A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in performance (for this measure, between providers). A value of 0.7 is often regarded as a minimum acceptable reliability value.
- A sample 1,724 (41.8%) of 4,124 physicians reported the measure Results showed the average number of quality reporting events for physicians included is 61.0 and reliability was 0.92 (high reliability). For the program required minimum of 10 quality reporting events, the reliability was 0.65 (moderate reliability).

- A missing data assessment was not performed. The developer states data missing from denominator excludes the patient from the measure, while missing numerator data counts as a measure "fail".

eMeasure Testing

- Critical data element (Validity against the Gold Standard) was conducted. Per NQF criteria, if empirical validity testing was performed of patient-level data, the rating from validity testing of patient-level data elements should be used. The developer provides simple agreement results for critical data element validity. Comparison of the values for several data elements (electronic extracted vs. data abstracted) using EHR data was conducted and could satisfy the data element reliability criterion (Algorithm box 3; validity testing results described in 2b.2 below). Results were provided for most, but not all, critical data elements. Percent agreement statistics were presented; however, percent agreement does not adjust for agreement due to chance, and should not be used alone to demonstrate reliability.
- Ideally, implementation of an eMeasure can be considered an automated process, and therefore the calculations will be consistent.
- In addition to critical data element testing, the developer submitted pre-testing from the Measure Authoring Tool within the Bonnie Output that tests eMeasure logic and performance calculation. This testing does use "live" EHR patients, though NQF currently accepts Bonnie Output pre-testing when EHR testing was not provided. Results in the 55 "pre-test" patients demonstrated 100% performance in identifying both expected and actual initial patient population, denominator, denominator exclusions, numerator, and denominator exceptions, with 82% of all possible data elements covered in the pre-test sample.
- For both specifications, a missing data assessment was not performed. The developer states data missing from denominator excludes the patient from the measure, while missing numerator data counts as a measure "fail".

Questions for the Committee (as appropriate):

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- Exclusions for arrhythmia, hypotension, asthma, bradycardia, atrioventricular block and cardiac pacer in Situ as exclusions are included in the eMeasure specification, but not in the Registry specifications.
- The evidence indicates that Acute Coronary Syndrome (ACS) is also appropriate for beta-blockades, which seem to appear in the provided value set conditions though not in the measure specifications.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

GPRO Registry Testing

- The developer provided face validity results using 12 members of the PCPI Measure Advisory Committee (MAC) with a mean rating of 4.17 and 91.7% of respondents either agree or strongly agree the measure is able to distinguish good and poor quality using a 1-5 Likert, highest score was 5. The MAC is independent of the measure developer experts.

eMeasure Testing

- The developer provided critical data element (Validity against the Gold Standard) was conducted. Per NQF criteria, if empirical validity testing was performed of patient-level data, this rating will also be used for reliability testing. Data elements included patient age, visit, problem list, or medical history of CAD, those who met the denominator

population.

- The developers provide simple agreement for the 134 patients sampled via automated EHR review. Of the sample, 111 patients (82.8%) were detected that met the numerator criteria. Performance on the measure was calculated to be 90.3% through comparison of automated and manual EHR review. When manual abstraction was added to the automated EHR review, agreement increased to 124 patients detected (92.5%). Additional statistical testing was not provided.
- Data element validity testing was conducting by comparing, for several data elements, the values obtained from electronic extraction from 1 EHR to those obtained from those abstracted from the EHR by an abstractor. Simple agreement was provided for most, but not all, critical data elements. However, simple agreement does not adjust for agreement due to chance and thus should not be used alone to demonstrate validity; sensitivity/specificity statistics are preferred for demonstrating data element validity. Percentage agreement values were relatively high for most data elements considered. It appears that only one abstractor was utilized, which is acceptable for testing validity against the gold standard in an EHR.
- In addition to critical data element testing, the developer submitted pre-testing from the Measure Authoring Tool within the Bonnie Output that also tests eMeasure performance calculation. This testing does use "live" EHR patients, though NQF currently accepts Bonnie Output pre-testing when EHR testing was not provided. Results in the 55 "pre-test" patients demonstrated 100% agreement for identifying both expected and actual initial patient population, denominator, denominator exclusions, numerator, and denominator exceptions. Testing characteristics are provided for the 55 "pre-test" patients, with 82% of the of data elements concepts included in the Initial Patient Population (IPP), Denominator, Numerator and Denominator Exceptions.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions/Exceptions:

GPRO Registry Testing

- Broad patient, medical & system reasons quality codes are reported for the Registry specification, with 4,291 exceptions reported from 1,724 physicians, with 2.5 exceptions reported per physician, and an overall exception rate of 3.9%. The timeframe for the frequencies and variability across providers is not provided.

eMeasure Testing

- An exception analysis of 2,717 exceptions show 2,292 (84.4%) exceptions were medical reasons for not prescribing beta blocker therapy, 347 (12.8%) exceptions were patient reasons and 78 (2.9%) were system reasons.
- Exclusions for arrhythmia, hypotension, asthma, bradycardia, atrioventricular block and cardiac pacer in Situ as exclusions are included in the eMeasure specification, but not in the Registry specifications. These do not appear in the exception analysis.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This measure is not risk adjusted.

2b5. Meaningful difference:

- Meaningful difference data for the Registry measure is provided.
- Meaningful difference data for the eMeasure was not provided by the developer, as the eMeasure testing data used fictitious "pre-test" patients, though higher performance and detection rates were noted in eMeasure validity testing with EHR & manual reviews, than in EHR-only reviews.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- NQF criteria require a comparability assessment of data sources/methods, such as with multiple specifications for the same measure. This was not provided by the developer, as the eMeasure testing data used fictitious "pre-test" patients.
- For validity testing of the eMeasure, developers did find higher performance & detection rates for the records abstracted with EHR & manual reviews, than EHR-only reviews.

2b7. Missing Data

- A missing data assessment was not performed.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- Measure specifications meet criteria and seem straightforward

2a2.: ***Committee's Comments on Reliability-Testing:***

- Signal to noice analysis is said to have a minimal acceptable reliability value of 0.7 and with a minimum of 10 reporting events, the reliability is only 0.65, but overall was 0.92.

2b1.: ***Committee's Comments on Validity-Specifications***:

- ACS is not included, rather just AMI, but I am fine with that.

2b2.: ***Committee's Comments on Validity-Testing:***

- Results in the 55 "pre-test" patients demonstrated 100% agreement for identifying both expected and actual initial patient population, denominator, denominator exclusions, numerator, and denominator exceptions. Testing characteristics are provided for the 55 "pre-test" patients, with 82% of the of data elements concepts included in the Initial Patient Population (IPP), Denominator, Numerator and Denominator Exceptions.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Patient level does not get assessed in this e-measure so and it is tested with limited numbers of "pre-test" patients so it is hard to assess meaningful difference or comparability of data sources.

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

GPRO Registry Testing

- For both Registry & eMeasures, data elements are collected by and used by healthcare personnel during provisions of care (BP, lab values), codes by someone other than the person obtaining the information (billing), and abstracted by someone other personnel (quality staff). All data elements are in electronic fields.
- The measure is specified for Clinician : Group/Practice, Clinician : Individual use.
- In the eMeasure Feasibility Scorecard, the developer states all data elements score "3" on a scale of 1 – 3 (3 the highest) for current and future use. The developer should clarify if this includes ICD-10, SNOMED-CT & RxNorm codes

for the eMeasure for all data elements. The EHR product(s) & number of EHRs used in the eMeasure Feasibility Scorecard is not reported. NQF requires a Feasibility Scorecard from more than one EHR.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- There was not a feasibility scorecard from more than one EHR reported, as per NQF requirements.

Criterion 4: Usability and Use

4. Usability and 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.

- The developer provides the following information on measure use
  - Public Reporting - Physician Quality Reporting System (PQRS)
  - Payment Program - Meaningful Use Stage 2 (EHR Incentive Program)
  - Quality Improvement with Benchmarking (external benchmarking to multiple organizations)- Pinnacle Registry
- The developer states no unintended consequences have been noted, though they monitor for identification and mitigation continuously.
- In 2014 the Measure Applications Partnership (MAP) Clinician Workgroup did not support the measure for the Physician Compare and Value-Based Payment Modifier Programs because the measure does not adequately address any current needs of the program preferring other outcome measures that address coronary artery disease. In 2015, the Clinician Workgroup did not support narrow-focused, process measures for MSSP.

Questions for the Committee:

- *Is the measure publicly reported?*
- *For maintenance measures – is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- Physician compare, value based modifier program and MSSP all do not support the use of this measure.

Criterion 5: Related and Competing Measures

- 0071: Persistence of Beta-Blocker Treatment After a Heart Attack
- 0083: Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD).

The developer states the specifications are harmonized to the extent possible. As a result, the denominator specifications for the measures differ where needed based on the differing patient populations.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0070

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

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: 6/30/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Beta-Blocker Therapy for CAD patients with Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

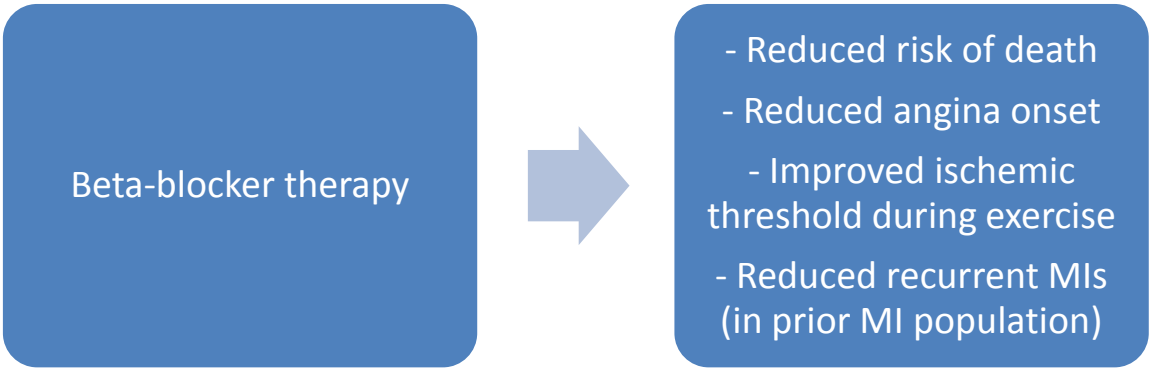
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Fihn SD, Gardin JM, Abrams J, et al. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2012;60:e44-164.

Available at: http://content.onlinejacc.org/article.aspx?articleid=1391404

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

4.4.2.2. BETA-BLOCKER THERAPY:

Beta-blocker therapy should be started and continued for 3 years in all patients with normal LV function after MI or ACS. (Class I, Level of Evidence: B)

Beta-blocker therapy should be used in all patients with LV systolic dysfunction (EF ≤40%) with heart failure or prior MI, unless contraindicated. (Use should be limited to carvedilol, metoprolol succinate,

or bisoprolol, which have been shown to reduce risk of death.) (Class I, Level of Evidence: A)

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Both recommendation statements included in section 1a.4.2 have been assigned a Class I recommendation. Class I recommendations refer to " Conditions for which there is evidence and/or general agreement that a given procedure or treatment is beneficial, useful, and effective."

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

ACCF/AHA guideline methodology categorizes indications as class I, II, or III on the basis of a multifactorial assessment of risk and expected efficacy viewed in the context of current knowledge and the relative strength of this knowledge. These classes summarize the recommendations for procedures or treatments as follows and noted in the table below:

Class I: Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective.

Class II: Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment.

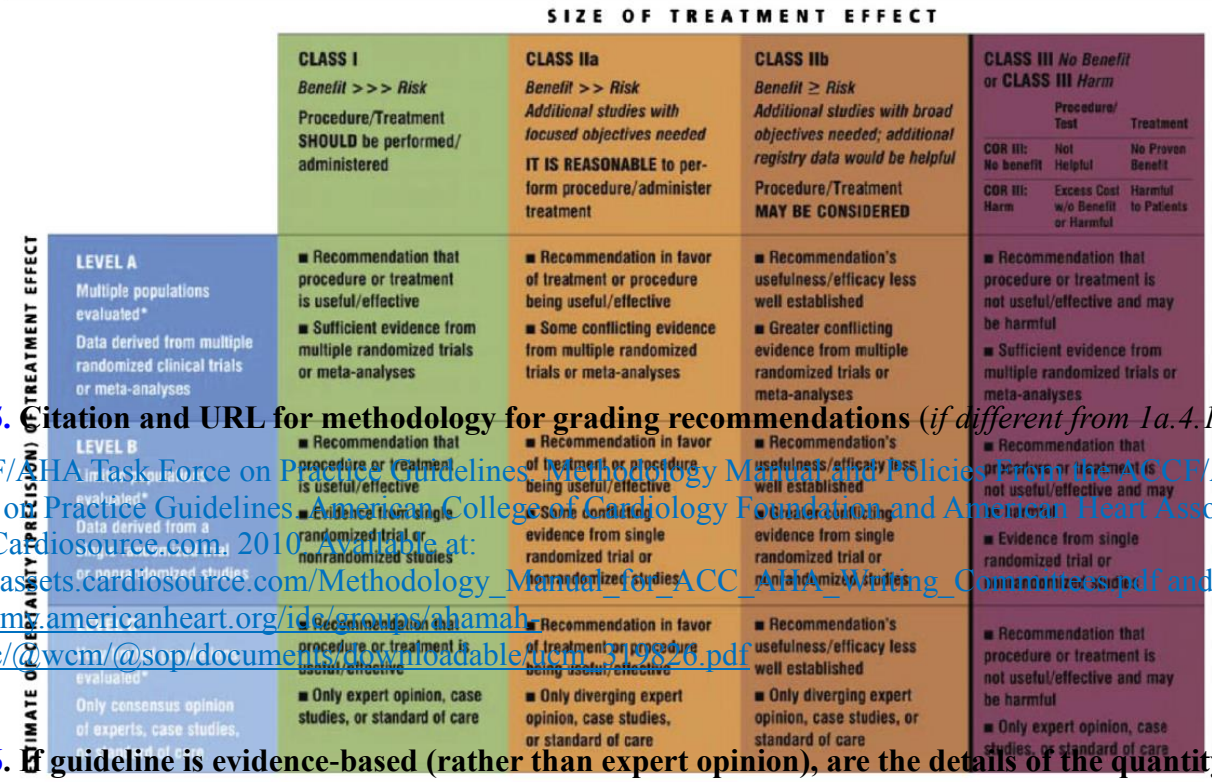
- IIa: Weight of evidence/opinion is in favor of usefulness/efficacy
- IIb: Usefulness/efficacy is less well established by evidence/opinion.

Class III: Conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective e and in some cases may be harmful.

- No Benefit- Procedure/Test not helpful or Treatment w/o established proven benefit
- Harm- Procedure/Test leads to excess cost w/o benefit or is harmful, and or Treatment is harmful

Additional detail regarding the classification of recommendation and level of evidence is provided in the following table.

Table 1. Applying Classification of Recommendation and Level of Evidence



1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

The section of the guideline which includes the recommendations referenced in 1a4.2. pertains to the use of beta-blocker therapy in patients with stable ischemic heart disease

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

The guideline recommendations refer to 2 distinct patient populations address by the measure – 1) patients with a prior (resolved) (within the past 3 years) myocardial infarction and 2) patients with left ventricular systolic dysfunction (LVEF <40%). For the prior MI population, the weight of the evidence in support of the recommendation is rated as Level B. Level B evidence refers to "Data derived from a single randomized trial, or nonrandomized studies." For the LVSD population, the weight of the evidence in support of the recommendation is rated as Level A. Level A evidence refers to "Data derived from multiple randomized clinical trials or meta-analyses."

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Levels A and B are described in 1a.7.2. Level C evidence refers to "Only consensus opinion of experts, case studies, or standard-of-care." Additional details and information about the evidence rating scheme can also be seen in 1a.4.2. and 1a.4.3.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 1996-2009

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Information regarding the total number of studies and type of study designs included in the body of evidence is not available.

However, for the prior MI population: the guideline cites 3 articles in support of the recommendation statement. They include 2 systematic reviews including 33 and 82 randomized controlled trials, respectively, dating back to 1980. The third article was an observational study.

For the LVSD population: the guideline cites 5 articles in support of the recommendation statement. They include 3 randomized controlled trials, 1 meta-analysis of randomized controlled trials and 1 comparative analysis of randomized controlled trials.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Information regarding the overall quality of evidence across studies is not available.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The guideline does not include an overall estimate of benefit from the body of evidence. However, they do include the following summary information regarding the benefits of beta-blocker therapy, "Decreases in the rate–BP product, AV nodal conduction, and myocardial contractility from beta blockers reduce myocardial oxygen demand, counteracting beta-receptor activity and contributing to a reduction in angina onset, with improvement in the ischemic threshold during exercise and in symptoms. These agents significantly reduce deaths and recurrent MIs in patients who have suffered a MI and are especially effective when a STEMI is complicated by persistent or recurrent ischemia or tachyarrhythmias early after the onset of infarction."

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

The guideline describes the principle adverse effects of beta blockers as fatigue, exercise intolerance, lethargy, insomnia, nightmares, and impotence.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

The guidelines reviewed and incorporated relevant new clinical trials published in peer-reviewed journals and articles through December 2011. A Medical Subject Headings (MeSH®) was conducted using the terms "Adrenergic beta-Antagonists"[Mesh] AND "Coronary Artery Disease"[Mesh] to identify articles published after 2011, resulting in 75 articles.

The articles that are most relevant to the focus of the body of evidence are described below.

1. Citation: Bangalore S1, Steg G, Deedwania P, Crowley K, Eagle KA, Goto S, Ohman EM, Cannon CP, Smith SC, Zeymer U, Hoffman EB, Messerli FH, Bhatt DL; REACH Registry Investigators. β-Blocker use and clinical outcomes in stable outpatients with and without coronary artery disease. JAMA. 2012 Oct 3;308(13):1340-9. doi: 10.1001/jama.2012.12559.

Description: Longitudinal, observational study of patients in the Reduction of Atherothrombosis for Continued Health (REACH) registry who were divided into 3 cohorts: known prior MI (n = 14,043), known CAD without MI (n = 12,012), or those with CAD risk factors only (n = 18,653) to assess the association of β-blocker use with cardiovascular events in stable patients with a prior history of MI, in those with CAD but no history of MI, and in those with only risk factors for CAD.

Results: With a median follow-up of 44 months (interquartile range, 35-45 months), event rates were not significantly different in patients with β-blocker use compared with those without β-blocker use for any of the outcomes tested, even in the prior MI cohort (489 [16.93%] vs 532 [18.60%], respectively; hazard ratio [HR], 0.90 [95% CI, 0.79-1.03]; P = .14). In the CAD without MI cohort, the associated event rates were not significantly different in those with β-blocker use for the primary outcome (391 [12.94%]) vs without β-blocker use (405 [13.55%]) (HR, 0.92 [95% CI, 0.79-1.08]; P = .31), with higher rates for the secondary outcome (1101 [30.59%] vs 1002 [27.84%]; odds ratio [OR], 1.14 [95% CI, 1.03-1.27]; P = .01) and for the tertiary outcome of hospitalization (870 [24.17%] vs 773 [21.48%]; OR, 1.17 [95% CI, 1.04-1.30]; P = .01). In the cohort with CAD risk factors only, the event rates were higher for the primary outcome with β-blocker use (467 [14.22%]) vs without β-blocker use (403 [12.11%]) (HR, 1.18 [95% CI, 1.02-1.36]; P = .02), for the secondary outcome (870 [22.01%] vs 797 [20.17%]; OR, 1.12 [95% CI, 1.00-1.24]; P = .04) but not for the tertiary outcomes of MI (89 [2.82%] vs 68 [2.00%]; HR, 1.36 [95% CI, 0.97-1.90]; P = .08) and stroke (210 [6.55%] vs 168 [5.12%]; HR, 1.22 [95% CI, 0.99-1.52]; P = .06). However, in those with recent MI (≤1 year), β-blocker use was associated with a lower incidence of the secondary outcome (OR, 0.77 [95% CI, 0.64-0.92]).

Conclusion: Although this observational study found that the use of β-blockers in the populations studied was not associated with a lower risk of composite cardiovascular events, the article received several letters which highlighted 2 primary concerns: 1) the use of an observational study to assess the effectiveness of a drug when large RCTs and meta-analyses already have shown its effectiveness and 2) the study did not distinguish among different types of beta-blockers. 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 the study taking into account the full body of evidence available.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*
For patients with coronary artery disease (CAD), beta-blockers are recommended for 3 years after myocardial infarction or acute coronary syndrome. Beta-blockers, particularly carvedilol, metoprolol succinate, or bisoprolol which have been shown to reduce risk of death, are recommended indefinitely for patients with CAD and LV systolic dysfunction. These agents have proven efficacy in reducing angina onset and improving the ischemic threshold during exercise. In patients who have suffered an MI, beta-blockers significantly reduce deaths and recurrent MIs. (1) Nonadherence to cardioprotective medications is prevalent among outpatients with CAD and can be associated with a broad range of adverse outcomes, including all-cause and cardiovascular mortality, cardiovascular hospitalizations, and the need for revascularization procedures.(2) This measure is intended to promote beta-blocker usage in select patients with CAD.

References:
1. Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, Douglas PS, Foody JM, Gerber TC, Hinderliter AL, King SB III, Kligfield PD, Krumholz HM, Kwong RYK, Lim MJ, Linderbaum JA, Mack MJ, Munger MA, Prager RL, Sabik JF, Shaw LJ, Sikkema JD, Smith CR Jr, Smith SC Jr, Spertus JA, Williams SV. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2012;60:e44-164.
2. Gibbons RJ, Abrams J, Chatterjee K, Daley J, Deedwania PC, Douglas JS, Ferguson TB Jr., Fihn SD, Fraker TD Jr., Gardin JM, O'Rourke RA, Pasternak RC, Williams SV. ACC/AHA 2002 guideline update for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1999 Guidelines for the Management of Patients with Chronic Stable Angina). 2002. Available at: www.acc.org/clinical/guidelines/stable/stable.pdf

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
2013 PQRS Experience Report
2013 is the most recent year for which PQRS Experience Report measure data is available. The average performance rates on Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%) over the last several years are as follows:

2010: 71.4%
2011: 82.1%
2012: 69.9%
2013: 74.2%

2013 Small Group Practice Exception Rate: 2.0%

It is important to note that PQRS was a voluntary reporting program, with approximately 51% of eligible professionals participating using any reporting option in 2013, and performance rates may not be nationally representative.
Reference: Center for Medicare and Medicaid Services. 2013 Reporting Experience Including Trends (2007-2014). Available: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/AnalysisAndPayment.html

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.
Suboptimal rates of beta-blocker prescriptions among patients with CAD indicated by PQRS data are further evidenced by several recent studies.

Maddox and colleagues analyzed data from 2008 through 2010 from the NCDR® PINNACLE Registry®, a national outpatient cardiology practice registry, to assess practice variation of secondary prevention medication prescription among CAD patients. Among eligible patients, beta-blockers were prescribed in 73.3% (63,800/86,999) at their index clinic visit. After inclusion of all visits among eligible patients occurring within the year following the index visit, the rates increased to 77.3%. Among practices, the median prescription rate of beta-blockers for eligible patients at their index clinic visit was 78.4% (range 35.2-100%) and 79.4% (range 46.2-100%) after inclusion of all visits among eligible patients occurring within the year following the index visit.(1)

An earlier study by Chan and colleagues analyzed 2008-9 data from the Pinnacle registry and found slightly higher rates (86.4%) of beta-blocker prescription among CAD patients following an MI. It's important to note that the Chan et al. study examined compliance rates with performance measures among the first 14,000 outpatients enrolled in the PINNACE program as compared to the Maddox et al study which included a larger and more heterogeneous patient and practice population.(2)

References:
1. Maddox TM, Chan PS, Spertus JA, Tang F, Jones P, Ho PM, Bradley SM, Tsai TT, Bhatt DL, Peterson PN. Variations in coronary artery disease secondary prevention prescriptions among outpatient cardiology practices: insights from the NCDR (National Cardiovascular Data Registry). J Am Coll Cardiol. 2014 Feb 18;63(6):539-46. doi: 10.1016/j.jacc.2013.09.053. Epub 2013 Oct 30.

2. Chan PS, Oetgen WJ, Buchanan D, et al. Cardiac performance measure compliance in outpatients: the American College of Cardiology and National Cardiovascular Data Registry´s PINNACLE (Practice Innovation And Clinical Excellence) program. J. Am. Coll. Cardiol. 2010; 56(1):8–14.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) 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 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
The Chan et al. article cited above conducted a secondary analysis of PINNACLE data for select performance measures to examine whether compliance rates differed by race or sex. The authors found that compliance rates were similar between black and white patients and men and women for all 4 CAD performance measures (including beta-blocker therapy after MI). (1)

A separate analysis was completed using PINNACLE data from 2009 to compare treatment rates by insurance status for 5 quality-of-care indicators for CAD care related to medication treatment. Uninsured patients were less likely to receive ß-blocker therapy after MI as compared with those who had private health insurance (73.3% vs. 80.5%; unadjusted RR=0.91; 95% CI, 0.87-0.95; P<0.001). There were no meaningful differences in treatment rates between patients with public and private insurance. (2)

1. Chan PS, Oetgen WJ, Buchanan D, et al. Cardiac performance measure compliance in outpatients: the American College of Cardiology and National Cardiovascular Data Registry´s PINNACLE (Practice Innovation And Clinical Excellence) program. J. Am. Coll. Cardiol. 2010; 56(1):8–14.

2. Smolderen KG, Spertus JA, Tang F, et al. Treatment Differences by Health Insurance Among Outpatients with Coronary Artery Disease: Insights from the NCDR®. J Am Coll Cardiol. 2013 Mar 12; 61(10): 1069–1075.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Heart disease is the leading cause of death for both men and women in the United States and coronary heart disease is the most common type of heart disease.(1) According to the American Heart Association's 2015 Heart Disease and Stroke Statistics, coronary heart disease alone caused ˜1 of every 7 deaths in the United States in 2011. In 2011, close to 400,000 Americans died of coronary heart disease. Each year, an estimated ˜635 000 Americans have a new coronary attack (defined as first hospitalized myocardial infarction or coronary heart disease death) and ˜300 000 have a recurrent attack. It is estimated that an additional 155 000 silent first myocardial infarctions occur each year. Approximately every 34 seconds, 1 American has a coronary event, and approximately every 1 minute 24 seconds, an American will die of one.(2)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. CDC, NCHS. Underlying Cause of Death 1999-2013 on CDC WONDER Online Database, released 2015. Data are from the Multiple Cause of Death Files, 1999-2013, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.
2. Mozaffarian D, Benjamin EJ, Go AS, et al.; on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation. 2015;131:e29–e322.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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 : Ischemic Heart Disease, Coronary Artery Disease

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

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 included as an attachment with this submission. Additional measure details may be found at:http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/eCQM_Library.html Value sets at 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:** EP_CMS145v4_NQF0070_CAD_BB.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:** 0070_AMAPCPI_CAD-BB_ValueSets_June2015.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Supporting guidelines and coding value sets included in the measure are reviewed on an annual basis. The measure and specifications have been updated to align with the 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the Diagnosis and Management of Patients With Stable Ischemic Heart Disease which now recommends the use of beta-blocker therapy for 3 years following MI or ACS as opposed to indefinite use as previously recommended. Additional limited changes have been incorporated during the annual update process to adhere to current eCQM industry standards while preserving the original measure intent.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who were prescribed beta-blocker therapy

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Once during the 12 consecutive month measurement period

S.6. **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, 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.
For EHR:
HQMF eMeasure developed and is included in this submission.

For Registry:
Option 1 – for patients with LVEF < 40%:
Definitions:
Prescribed- May include 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.
Beta-blocker Therapy- For patients with prior LVEF < 40%, beta-blocker therapy includes the following: bisoprolol, carvedilol, or sustained release metoprolol succinate.

Report Quality Data Code, G9189: Beta-blocker therapy prescribed or currently being taken

Option 2 – for patients with prior MI:
Definitions:

Prescribed- May include 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.
Beta-blocker Therapy- For patients with prior MI, beta-blocker therapy includes any agent within the beta-blocker drug class. As of 2014, no recommendations or evidence are cited in current stable ischemic heart disease guidelines for preferential use of specific agents.

Report CPT Category II Code, 4008F: Beta-blocker therapy prescribed or currently being taken

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who also have a prior MI (within the past 3 years) or a current or prior LVEF <40%

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
For EHR:
HQMF eMeasure developed and is included in this submission.

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

Prior Myocardial Infarction (MI) for denominator 2 is limited to those occurring within the past 3 years.

DENOMINATOR NOTES:
The requirement of "Count >=2 of Encounter, Performed" is to establish that the eligible professional has an existing relationship with the patient.

For Registry:
Option 1 -- for patients with LVEF < 40%:
Patient aged >= 18 years
AND
Diagnosis for coronary artery disease (ICD-9-CM) [for use 1/1/2015-9/30/2015]: 411.0, 411.1, 411.81, 411.89, 413.0, 413.1, 413.9, 414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.2, 414.3, 414.8, 414.9, V45.81, V45.82
Diagnosis for coronary artery disease (ICD-10-CM) [for use 10/01/2015-12/31/2015]: I20.0, I20.1, I20.8, I20.9, I24.0, I24.1, I24.8, I24.9, I25.10, I25.110, I25.111, I25.118, I25.119, I25.5, I25.6, I25.700, I25.701, I25.708, I25.709, I25.710, I25.711, I25.718, I25.719, I25.720, I25.721, I25.728, I25.729, I25.730, I25.731, I25.738, I25.739, I25.750, I25.751, I25.758, I25.759, I25.760, I25.761, I25.768, I25.769, I25.790, I25.791, I25.798, I25.799, I25.810, I25.811, I25.812, I25.82, I25.83, I25.89, I25.9, Z95.1, Z95.5, Z98.61
OR
History of cardiac surgery (CPT): 33140, 33510, 33511, 33512, 33513, 33514, 33516, 33517, 33518, 33519, 33521, 33522, 33523, 33533, 33534, 33535, 33536, 92920, 92924, 92928, 92933, 92937, 92941, 92943
AND
Patient encounter(s) during reporting period (CPT): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99304, 99305, 99306, 99307, 99308, 99309, 99310, 99324, 99325, 99326, 99327, 99328, 99334, 99335, 99336, 99337, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350
AND
Two Denominator Eligible Visits
AND
Left ventricular ejection fraction (LVEF) < 40%: G8694

Option 2 – for patients with prior MI:
Patient aged >= 18 years

AND
Diagnosis for coronary artery disease (ICD-9-CM) [for use 1/1/2015-9/30/2015]: 411.0, 411.1, 411.81, 411.89, 413.0, 413.1, 413.9, 414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.2, 414.3, 414.8, 414.9, V45.81, V45.82
Diagnosis for coronary artery disease (ICD-10-CM) [for use 10/01/2015-12/31/2015]: I20.0, I20.1, I20.8, I20.9, I24.0, I24.1, I24.8, I24.9, I25.10, I25.110, I25.111, I25.118, I25.119, I25.5, I25.6, I25.700, I25.701, I25.708, I25.709, I25.710, I25.711, I25.718, I25.719, I25.720, I25.721, I25.728, I25.729, I25.730, I25.731, I25.738, I25.739, I25.750, I25.751, I25.758, I25.759, I25.760, I25.761, I25.768, I25.769, I25.790, I25.791, I25.798, I25.799, I25.810, I25.811, I25.812, I25.82, I25.83, I25.89, I25.9, Z95.1, Z95.5, Z98.61
OR
History of cardiac surgery (CPT): 33140, 33510, 33511, 33512, 33513, 33514, 33516, 33517, 33518, 33519, 33521, 33522, 33523, 33533, 33534, 33535, 33536, 92920, 92924, 92928, 92933, 92937, 92941, 92943
AND
Diagnosis for myocardial infarction (ICD-9-CM) [for use 1/1/2015-9/30/2015]: 410.00, 410.01, 410.02, 410.10, 410.11, 410.12, 410.20, 410.21, 410.22, 410.30, 410.31, 410.32, 410.40, 410.41, 410.42, 410.50, 410.51, 410.52, 410.60, 410.61, 410.62, 410.70, 410.71, 410.72, 410.80, 410.81, 410.82, 410.90, 410.91, 410.92, 412
Diagnosis for myocardial infarction (ICD-10-CM) [for use 10/01/2015-12/31/2015]: I21.01, I21.02, I21.09, I21.11, I21.19, I21.21, I21.29, I21.3, I21.4, I22.0, I22.1, I22.2, I22.8, I22.9, I24.1, I25.2
AND
Patient encounter(s) during reporting period (CPT): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99304, 99305, 99306, 99307, 99308, 99309, 99310, 99324, 99325, 99326, 99327, 99328, 99334, 99335, 99336, 99337, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350
AND
Two Denominator Eligible Visits

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Documentation of medical reason(s) for not prescribing beta-blocker therapy (eg, 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 health care system)

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
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. This measure was developed using the PCPI exception methodology which 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 this measure, exceptions may include medical reason(s) (eg, allergy, intolerance, other medical reasons), patient reason(s) (eg, patient declined, other patient reasons) or system reason(s) for not prescribing beta-blocker therapy (eg, other reasons attributable to the health care system).

Where examples of exceptions are included in the measure language, value sets for these examples are developed and included in the eSpecifications. 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.

Additional details by data source are as follows:
For EHR:
HQMF eMeasure developed and is included in this submission.

For Registry:

Option 1 -- for patients with LVEF < 40%:
Report Quality Data Code, G9190: Documentation of medical reason(s) for not prescribing beta-blocker therapy (eg, allergy, intolerance, other medical reasons)

Report Quality Data Code, G9191: Documentation of patient reason(s) for not prescribing beta-blocker therapy (eg, patient declined, other patient reasons)

Report Quality Data Code, G9192 : Documentation of system reason(s) for not prescribing beta-blocker therapy (eg, other reasons attributable to the health care system)

Option 2 – for patients with prior MI:
Append a modifier to CPT Category II Code:
4008F-1P : Documentation of medical reason(s) for not prescribing beta-blocker therapy (eg, allergy, intolerance, other medical reasons)

4008F-2P : Documentation of patient reason(s) for not prescribing beta-blocker therapy (eg, patient declined, other patient reasons)

4008F-3P : Documentation of system reason(s) for not prescribing beta-blocker therapy (eg, other reasons attributable to the health care system)

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 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.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
No risk adjustment or risk stratification

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
No risk adjustment or risk stratification

S.16. Type of score:
Rate/proportion
If other:

S.17. 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.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*
To calculate performance rates:

1. Find the patients who meet the initial population (ie, 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 (ie, 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 (ie, 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) (eg, allergy, intolerance, other medical reasons), patient reason(s) (eg, patient declined, other patient reasons) or system reason(s) for not prescribing beta-blocker therapy (eg, other reasons attributable to the health care system).] 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 (ie, 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.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable. The measure is not based on a sample.

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. The measure is not based on a survey.

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Patient eligibility is determined by a set of defined criteria relevant to a particular measure. If data required to determine patient eligibility are missing, those patients/cases would be ineligible for inclusion in the denominator and therefore the patient/case would be deleted.

If data required to determine if a denominator eligible patient qualifies for the numerator (or has a valid exclusion/exception) are missing, this case would represent a quality failure.

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry

S.24. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Not applicable.

S.25. 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.26. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)* Ambulatory Care : Clinician Office/Clinic, Home Health, Other, Post Acute/Long Term Care Facility : Long Term Acute Care Hospital, Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility If other: Domiciliary
S.28. **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.)*
2a. Reliability – See attached Measure Testing Submission Form **2b. Validity – See attached Measure Testing Submission Form** NQF_0070_Coronary_Artery_Disease_-CAD-_-_Beta_Blocker_Therapy_Prior_to_MI_or_LVSD_Testing_Attachment.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: CoronaryArtery Disease – Beta Blocker Therapy Prior to MI or LVSD
Date of Submission: Click here to enter a date

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☑ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specifications** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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.

2b2. 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.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

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.

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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☑ clinical database/registry	☑ clinical database/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*).

Data 1 (EHR - Validity Against the Gold Standard)

The data source is EHR data.

Bonnie Patient Test Deck

As a supplement to the EHR reliability testing performed on this measure, a deck of patient test cases have 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 2 (GPRO Registry)

The data source is the Centers for Medicare & Medicaid Service (CMS) PQRS GPRO database.

Data 3 (EHR – Exceptions Analysis)

The data source is EHR data.

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

Data 1 (EHR - Validity Against the Gold Standard)

The data are collected from patients sampled from 2004.

Data 2 (GPRO Registry)

The data are for the time period January 2013 – December 31, 2013 and cover the entire United States.

Data 3 (EHR – Exceptions Analysis)

The data are collected from patients sampled in 2009.

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.26***)	**Measure Tested at Level of:**
☑ individual clinician	☑ individual clinician
☑ group/practice	☑ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the*

analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample)

Data 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. The clinic employs 40 full or part-time internal medicine physicians and provides more than 41,000 patient visits annually.

Data 2 (GPRO Registry)

The total number of physicians reporting on this measure is 4,124. Of those, 1,724 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, 41.8 percent of physicians are included in the analysis, and the average number of quality reporting events is 61.0 for a total of 105,124 events. The range of quality reporting events for 1,724 physicians included is from 1398 to 10. The average number of quality reporting events for the remaining 58.2 percent of physicians who aren't included is 3.6.

Data 3 (EHR – Exceptions Analysis)

The data sampled came from 5 physician offices using 5 different EHR systems.

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

Data 1 (EHR - Validity Against the Gold Standard)

The sample consisted of approximately 134 charts for a total of 134 eligible patients. One trained investigator reviewed the 134 charts. The patients were selected using random sampling.

Data 2 (GPRO Registry)

There were 105,124 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 3 (EHR – Exceptions Analysis)

The sample consisted of approximately 2,717 eligible patients.

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

Data 1 (EHR - Validity Against the Gold Standard)

The data sample was used for the purposes of reliability and validity testing.

Data 2 (GPRO Registry)

The same data sample from each data source was used for reliability testing and exceptions analysis.

Face Validity (Data 2)

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 3 (EHR – Exceptions Analysis)

The data sample was used for the exception analysis only.

1.8. What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Data 1 (EHR - Validity Against the Gold Standard)

Patient-level sociodemographic (SDS) variables were not analyzed in this project.

Data 2 (GPRO Registry)

This was not captured as part of the testing.

Data 3 (EHR – Exceptions Analysis)

This was not captured as part of the testing.

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*)

Data 1 (EHR - Validity Against the Gold Standard)

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

Data 2 (GPRO Registry / 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.

2a2.3. For each level 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*)

Data 1 (EHR - Validity Against the Gold Standard)

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

Data 2 (GPRO Registry)

For this measure, the reliability at the minimum level of quality reporting events (10) was 0.65. The average number of quality reporting events for physicians included is 61.0. The reliability at the average number of quality reporting events was 0.92.

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?*)

Data 1 (EHR - Validity Against the Gold Standard)

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

Data 2 (GPRO Registry)

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.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. For each level 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)*

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 a visit, problem list, or medical history diagnosis of CAD.

Data 2 (GPRO Registry) Face Validity

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

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Data 1 (EHR - Validity Against the Gold Standard)

Of the 134 patients sampled via automated EHR review, 111 patients (82.8%) that met the numerator criteria were detected. Performance on the measure was calculated to be 90.3% through comparison of automated and manual EHR review.

Face Validity (Data 2)

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

2b2.4. 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?*)

Data 1 (EHR - Validity Against the Gold Standard)

Discrepancies between performance measures based on EHR automated review alone and those based on automated review plus manual reviews were due to two types of misclassification: failure to correctly identify performance of quality measures among true, eligible patients; and failure to correctly exclude patients. Upon further analysis, the differences between automated review alone and automated plus manual reviews were 10 patients (7.5%).

Face Validity (Data 2)

The results of the expert panel rating of the validity statement were as follows: N = 12; Mean rating = 4.17 and 91.7% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings

1 – 1 response (Strongly Disagree)

2 – 0 responses

3 – 0 responses (Neither Agree nor Disagree)

4 – 6 responses

5 – 5 responses (Strongly Agree)

2b3. EXCLUSIONS ANALYSIS

NA no exclusions — *skip to section 2b4*

2b3.1. 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*)

Data 1 (EHR - Validity Against the Gold Standard)

This data sample was not used to test exclusions.

Data 2 (GPRO Registry)

With the information available from the PQRS Registry, 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 3 (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.

2b3.2. 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*)

Data 1 (EHR - Validity Against the Gold Standard)

This data sample was not used to test exclusions.

Data 2 (GPRO Registry)

Amongst the 1,724 physicians with the minimum (10) number of quality reporting events, there were a total of 4,291 exceptions reported. The average number of exceptions per physician in this sample is 2.5. The overall exception rate is 3.9%.

Data 3 (EHR – Exceptions Analysis)

Review of the 2,717 exceptions revealed that 2,292 (84.4%) exceptions were medical reasons for not prescribing beta blocker therapy, 347 (12.8%) exceptions were patient reasons and 78 (2.9%) were system reasons.

2b3.3. 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*)

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.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

Not applicable

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. 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)*

Not applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

Not applicable

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. 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*)

Not applicable

***2b4.11. 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*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

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 Registry)

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

2b5.2. 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*)

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 Registry)

Based on the sample of 1,724 included physicians, the mean performance rate is 0.74, the median performance rate is 0.78 and the mode is 1.00. The standard deviation is 0.20. The range of the performance rate is 1.0, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.23 (0.65 - 0.88).

2b5.3. 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?*)

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 Registry)

The range of performance from 0.00 to 1.00 suggests there's clinically meaningful variation across physicians' performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

This test was not performed for this measure.

2b6.2. 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*)

This test was not performed for this measure.

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

This test was not performed for this measure.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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)

Data are not available to complete this testing.

2b7.2. 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)

Data are not available to complete this testing.

2b7.3. 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)

Data are not available to complete this testing.

3. Feasibility
Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
3a. Byproduct of Care Processes For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

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

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.**
Attachment **Attachment:** NQF_0070_Feasibility_Scorecard_Bonnie_Output_Screen_Shots_Revised.pdf

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
We have not identified any 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.

Planned	Current Use (for current use provide URL)
	Public Reporting Physician Quality Reporting System (PQRS) http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/MeasuresCodes.html Payment Program Meaningful Use Stage 2 (EHR Incentive Program) http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/eCQM_Library.html Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Pinnacle Registry http://cvquality.acc.org/en/NCDR-Home/Registries/Outpatient-Registries.aspx

4a.1. For each CURRENT use, checked above, provide:

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

1. Physician Quality Reporting System (PQRS) – Sponsored by the Centers for Medicare and Medicaid Services (CMS)
PQRS is a national reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs). The program provides an incentive payment to practices with EPs (identified on claims by their individual National Provider Identifier [NPI] and Tax Identification Number [TIN]). EPs satisfactorily report data on quality measures for covered Physician Fee Schedule (PFS) services furnished to Medicare Part B Fee-for-Service (FFS) beneficiaries (including Railroad Retirement Board and Medicare Secondary Payer). Beginning in 2015, the program also applies a payment adjustment to EPs who do not satisfactorily report data on quality measures for covered professional services. Source: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html. It is our understanding that CMS is also planning to move towards publicly reporting physician data via Physician Compare.

2. Meaningful Use Stage 2 (EHR Incentive Program) – Sponsored by the Centers for Medicare and Medicaid Services (CMS)
The Medicare and Medicaid EHR Incentive Programs provide incentive payments to eligible professionals, eligible hospitals, and critical access hospitals (CAHs) as they adopt, implement, upgrade or demonstrate meaningful use of certified EHR technology. Eligibility for incentive payments for the "meaningful use" of certified EHR technology is established if all program requirements are met, including successful implementation and reporting of program measures, which include this measure, to demonstrate meaningful use of EHR technology.

3. PINNACLE Registry (URL: http://cvquality.acc.org/en/NCDR-Home/Registries/Outpatient-Registries.aspx)
The PINNACLE Registry® is cardiology´s largest outpatient quality improvement registry, capturing data on coronary artery disease, hypertension, heart failure and atrial fibrillation. The PINNACLE Registry® continues to grow rapidly, with more than 2400 providers representing almost 800 unique office locations across the U.S submitting data to the registry as of the fourth quarter of 2013. As of the fourth quarter of 2013, the registry has more than 13 million patient encounter records. PINNACLE assists practices in understanding and improving care through the production and distribution of quarterly performance reports. These reports, covering all valid patient encounters, detail adherence to 28 cardiovascular clinical measures at the physician, location, and practice levels across coronary artery disease, hypertension, heart failure and atrial fibrillation. All jointly developed ACC/AHA/PCPI performance measures for these topics are reported by the registry.

4a.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 new insurance marketplace.

4a.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. 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.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

2013 PQRS Experience Report

2013 is the most recent year for which PQRS Experience Report measure data is available. The average performance rates on Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%) over the last several years are as follows:

2010: 71.4%
2011: 82.1%
2012: 69.9%
2013: 74.2%

2013 Small Group Practice Exception Rate: 2.0%

It is important to note that PQRS was a voluntary reporting program, with approximately 51% of eligible professionals participating using any reporting option in 2013, and performance rates may not be nationally representative.

Reference: Center for Medicare and Medicaid Services. 2013 Reporting Experience Including Trends (2007-2014). Available: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/AnalysisAndPayment.html

4b.2. **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 we create 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.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative**

unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.
We are not aware of any unintended consequences at this time, but we take unintended consequences very seriously and therefore continuously monitor to identify actions that can be taken to mitigate them.

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)**
0071 : Persistence of Beta-Blocker Treatment After a Heart Attack
0083 : Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measure 0070 addresses a patient population of patients with CAD and either a recent prior MI or LVSD. This patient population is also covered in part by the following NQF-endorsed measures: NQF 0071: Persistence of Beta-Blocker Treatment After a Heart Attack and NQF 0083: Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD). The specifications are harmonized to the extent possible. As a result, the denominator specifications for the measures differ where needed based on the differing patient populations.

5b. Competing Measures

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

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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or

methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** AMA-convened Physician Consortium for Performance Improvement
Co.2 **Point of Contact:** Samantha, Tierney, Samantha.Tierney@ama-assn.org, 312-464-5524-
Co.3 **Measure Developer if different from Measure Steward:** American Medical Association - Physician Consortium for Performance Improvement
Co.4 **Point of Contact:** Samantha, Tierney, Samantha.Tierney@ama-assn.org, 312-464-5524-

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.

Work Group members:
Joseph Drozda, MD, FACC (Co-Chair) (cardiology; methodology)
Joseph V. Messer, MD, MACC, FAHA (Co-Chair) (cardiology)
John Spertus, MD, FACC, FAHA (Co-Chair) (cardiology)
Bruce Abramowitz, MD, FACC (interventional cardiology; measure implementation)
Karen Alexander, MD, FACC (cardiology; geriatrics)
Craig T. Beam, CRE (patient representative)
Robert O. Bonow, MD, MACC, FAHA, FACP (cardiology)
Jill S. Burkiewicz, PharmD, BCPS (pharmacy)
Michael Crouch, MD, MSPH (family medicine)
David C. Goff, Jr., MD, PhD, FAHA, FACP (internal medicine)
Richard Hellman, MD, FACP, FACE (endocrinology)
Thomas James, III, FACP, FAAP (health plan representative)
Marjorie L. King, MD, FACC, MAACVPR (cardiology; cardiac rehabilitation)
Edison A. Machado, Jr., MD, MBA (measure implementation)
Eduardo Ortiz, MD, MPH (internal medicine; guideline development)
Michael O'Toole, MD, FACC (cardiology; electrophysiology; measure implementation)
Stephen D. Persell, MD, MPH (internal medicine; measure implementation)
Jesse M. Pines, MD, MBA, MSCE, FAAEM (emergency medicine)
Frank J. Rybicki, MD, PhD (radiology)
Lawrence B. Sadwin (patient representative)
Joanna D. Sikkema, MSN, ANP-BC, FAHA (cardiology)
Peter K. Smith, MD (thoracic surgery)
Patrick J. Torcson, MD, FACP, MMM (hospital medicine)
John B. Wong MD, FACP (internal medicine)

ACCF, AHA, and PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study are invited to participate as equal contributors to the measure development process. In addition, the ACCF, AHA and PCPI strive to include on their work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

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:** 12, 2014
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?** 12, 2015

Ad.6 **Copyright statement:** Copyright 2014 American College of Cardiology, American Heart Association and American Medical Association. All Rights Reserved.

Ad.7 **Disclaimers:** Physician Performance Measures (Measures) and related data specifications have been developed by the American Medical Association (AMA) - convened Physician Consortium for Performance Improvement(R) (PCPI[R]), American College of Cardiology (ACC) and American Heart Association (AHA). These Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Measures require a license agreement between the user and the AMA, (on behalf of the PCPI), ACC or AHA. Neither the AMA, ACC, AHA, PCPI nor its members shall be responsible for any use of the Measures.

THE MEASURES 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 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 2004-2014 American Medical Association. LOINC(R) copyright 2004-2014 Regenstrief Institute, Inc. This material contains SNOMED Clinical Terms(R) (SNOMED CT[R]) copyright 2004-2014 International Health Terminology Standards Development Organisation. ICD-10 copyright 2014 World Health Organization. All Rights Reserved.

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 0071
Measure Title: Persistence of Beta-Blocker Treatment After a Heart Attack
Measure Steward: National Committee for Quality Assurance
Brief Description of Measure: The percentage of patients 18 years of age and older during the measurement year who were hospitalized and discharged from July 1 of the year prior to the measurement year to June 30 of the measurement year with a diagnosis of acute myocardial infarction (AMI) and who received persistent beta-blocker treatment for six months after discharge.
Developer Rationale: This measure addresses the appropriate clinical management of a person who has experienced an AMI. Persistent beta-blocker treatment after a heart attack reduces the risk of mortality, reduces the risk and severity of reinfarction, and improves the preservation of the left ventricular function.

Numerator Statement: Patients who had a 180-day course of treatment with beta-blockers post discharge.
Denominator Statement: Patients 18 years of age and older as of December 31 of the measurement year who were hospitalized and discharged from July 1 of the year prior to the measurement year to June 30 of the measurement year with diagnosis of AMI. See question S.9 Denominator Details for methods to identify patients who qualify for the denominator.
Denominator Exclusions: Exclude from the denominator, hospitalizations in which the patient was transferred directly to a nonacute care facility for any diagnosis.

Exclude patients who are identified as having an intolerance or allergy to beta-blocker therapy. Any of the following anytime during the patient's history through the end of the continuous enrollment period meet criteria:
- Asthma (Asthma Value Set).
- COPD (COPD Value Set).
- Obstructive chronic bronchitis (Obstructive Chronic Bronchitis Value Set).
- Chronic respiratory conditions due to fumes and vapors (Chronic Respiratory Conditions Due to Fumes/Vapors Value Set).
- Hypotension, heart block >1 degree or sinus bradycardia (Beta-Blocker Contraindications Value Set).
- A medication dispensing event indicative of a history of asthma (Table PBH-D).
- Intolerance or allergy to beta-blocker therapy.

Measure Type: Intermediate Clinical Outcome
Data Source: Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy
Level of Analysis: Health Plan, Integrated Delivery System
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**
For a MAINTENANCE, what is the Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Jan 18, 2012

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

Public and Member Comments included:
- The facility can evaluate whether the patient has the resources to comply with medication recommendations and when available refer them to low-cost resources when they do not. The patient though is responsible for compliance. The facility and physicians can only control whether or not the beta-blocker treatment is prescribed.
- Support endorsement of this measure given a significant gap in performance.

Developer Response:
- The improvement in patient outcomes occurs only if the patients take the medication.

Clinicians can greatly influence patient compliance.
Steering Committee: Agree with developer's response.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for an *intermediate clinical 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. Per NQF's submission criteria, this medication adherence measure is an *intermediate clinical outcomes* measure, rather than a process measure as defined by the developer.

The developer provides the following evidence for this intermediate clinical outcome measure:

- This is a health plan/integrated delivery system intermediate clinical outcome measure that calculates the percentage of patients 18 years of age and older during the measurement year who were hospitalized and discharged from July 1 of the year prior to the measurement year to June 30 of the measurement year with a diagnosis of acute myocardial infarction (AMI) and who received persistent beta-blocker treatment for six months after discharge. The developer does not describe the methodology for determining the 75% threshold of 180.
- The developer provides 2 clinical practice guidelines with 3 guideline statements for the persistent use of beta-blockers in patients diagnosed with AMI. Grading is provided for each guideline statement including for beta blockades during and after hospitalization for a STEMI a Class I/Level B, for beta-blockades in HF patients after Non-STEMI/ACS hospitalization Class I/Level C, and beta-blockades for patients with normal LVF Class IIa/Level C. They also provided a 1966-1997 seminal systematic review summarizing the quantity, quality and consistency of the evidence, and supporting the recommendations with findings, with Levels B and C "Estimate of Certainty (Precision) of Treatment".
- The developer provides decision logic from secondary prevention to intermediate clinical outcome for the persistent use of beta-blockers in reducing the risk of mortality, risk and severity of re-infarction and improving the preservation of the left ventricular function with patients with AMI.

Questions for the Committee:
- *Is the evidence directly applicable to the process of care being measured?*
- *Is the process of care proximal and closely related to desired outcomes?*
- *Does the SC agree with the methodology for determining the 75% threshold for receiving beta blockades?*

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

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

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

  Commercial (HMO and PPO Combined)

YEAR	N	MEAN	ST DEV	10th	25th	50TH	75th	90th	IQ
2012	250	79%	8%	68%	73%	79%	84%	84%	11%
2013	252	81%	8%	71%	77%	82%	87%	91%	10%
2014	253	83%	7%	73%	78%	82%	88%	91%	10%

Medicaid

YEAR	N	MEAN	ST DEV	10th	25th	50TH	75th	90th	IQ
2012	54	80%	10%	67%	73%	83%	88%	91%	15%
2013	64	82%	9%	71%	78%	83%	88%	91%	10%
2014	75	84%	10%	72%	80%	86%	91%	95%	11%

Medicare

YEAR	N	MEAN	ST DEV	10th	25th	50TH	75th	90th	IQ
2012	255	87%	6%	79%	84%	88%	91%	94%	7%
2013	269	89%	6%	81%	85%	90%	93%	95%	8%
2014	269	90%	6%	83%	87%	91%	94%	96%	7%

N= # of health plans

- A review of the performance data provided appears to demonstrate similar commercial and Medicaid rates, and higher Medicare rates, with all rates increasing annually. Further explanation of the performance data characteristics and findings is not provided.
- The developer summarizes data from the literature on the prevalence of heart disease, medication adherence among MI survivors by disability, status, race/ethnicity, and income for all Medicare fee-for-service beneficiaries and the impact of employment status on rates of CHD/stroke.
- The developer states they do not collect performance data stratified by race, ethnicity, or language, though other HEDIS measure sets are available for stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - Moderate rating for evidence. QQC provided, with grades B and C quality for guidelines and systematic review. Additional, more recent evidence supports SR. Evidence applies directly to process.
   - This is an intermediate clinical outcome measure regarding prescription of beta blockers for 180 days post discharge from acute MI. the rationale is that persistent beta blocker treatment after an MI will reduce the risk of mortality, reduce the risk and severity of reinfarction and improve the preservation of LV function The evidence provided is from two clinical guidelines: one on management of STEMI and one on management of NONSTEMI. The STEMI guideline is graded level B that Beta blockers should be continued during and after hospitalization for all patients...but does not discuss reinfarction rates or outcomes. The actual guideline says the benefit of beta blockers for secondary prevention has been established in numerous trials conducted in the prereperfusion era and appears to be greatest for patients with MI complicated by HF, LV dysfunction or ventricular arrhythmias....long term duration not been prospectively addressed. Do NOT feel evidence supports the measure NONSTEMI guidelines...graded as level C THEREFORE would rate as Insufficient Evidence with exception

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- Moderate rating for evidence. QQC provided, with grades B and C quality for guidelines and systematic review.

Additional, more recent evidence supports SR. Evidence applies directly to process.
- Other systematic review: FREEMANTLE systematic review is old 1999 Insufficient evidence with exception

1b. ***Committee's Comments on Performance Gap:***
- Minimal performance gap noted, particularly in Medicare population. No disparities related to measure noted, though some data on indirect proxies.
- There is no statistical data to demonstrate gap in care but there is evidence that there is a disparities issue. This appears to be a disparities sensitive measure

1c. ***Committee's Comments on Composite Performance Measure:***
- Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.
- Data sources include administrative claims and electronic clinical data including pharmacy claims.
- ICD-9, ICD-10, UBREV, and UBTOB (type of billing) codes are provided for the numerator and denominator that is readily available from administrative claims data.
- "Persistent" beta-blocker treatment is defined as at least 75% (135 days) of six months (180 days) of filled prescriptions post hospital discharge. The developer permits "allowed gap days", though further explanation is not provided.
- Beta-blockers used to calculate the numerator are provided.
- Definitions for transfers and readmissions are provided.
- Patients who are direct transfers to non-acute facilities are excluded, as well as those with intolerance and/or allergy to beta-blocker excluded (e.g. asthma, COPD, hypotension); value sets provided including medications used to identify asthma. Patients excluded for asthma are defined with diagnostic and medication codes. This structure is not similarly utilized for other pulmonary conditions in the exclusions.
- Changes to the measure since last endorsement are not provided, rather categorized as "not significant" by the developer.

Questions for the Committee:
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- Reliability testing for this measure was conducted at the level of the performance measure score, using the data source and level of analysis specified. The primary analysis was conducted at the health plan level and stratified by product line (253 Commercial plans, 75 Medicaid plans and 269 Medicare plans) and includes all health plans submitting data to NCQA for HEDIS in 2013.
- A signal-to-noise analysis using the beta-binomial model was conducted. This type of analysis, which is appropriate for the measure, quantifies the amount of variation in performance that is due to differences between providers (as opposed to differences that are due to random measurement error). The method results

in a reliability statistic for each health plan.

- Overall reliability (likely the average reliability for plans in each stratum) for this measure is based on data submitted from 253 commercial plans, 75 Medicaid plans, and 269 Medicare plans is between 0.78 and 0.81. A reliability of 0.70 is generally considered a minimum threshold for acceptability.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The clinical practice guidelines supporting this measure recommend the persistent use of beta-blockers in patients with AMI.
- "Persistent" beta-blocker treatment is defined as at least 75% (135 days) of six months (180 days) of filled prescriptions post hospital discharge. The timeframe for determining "persistent" use and "allowed gap days", are not provided.
- The developers outline a seven step ICD-9 to ICD-10 conversion process.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- Construct validity was tested by exploring whether Persistence of Beta-Blocker Treatment after a Heart Attack (#0071) correlated with 3 medication measures using a Person correlation test. The developers used the same reliability testing data source for validity testing. Developers hypothesize that plans that do well in managing patients with AMI will also do well in managing patients with other conditions. Results of the correlation analysis suggest that results for this measure are moderately and statistically significantly correlated with results from the diabetes and cardiovascular disease cholesterol control measures and COPD medication measure, confirming the construct validity hypotheses.
- Systematic assessment of face validity consists of a six step standardized process called the HEDIS measure life cycle. The steps include identifying areas of interest/gaps in care, a literature review, measure development, public comment, data collection and ongoing evaluation.
- The developers state this measure was tested for face validity with three expert panels, Cardiovascular Measurement Advisory Panel (8 physicians, 1 nurse), Technical measurement Advisory Panel (12 members) and NCQA's Clinical Programs Committee (17 members). The developer does not provide statistical results from validity testing but does state that the multiple NCQA panels agreed that the measure, as specified, accurately assesses persistence of beta blocker use for at-risk patients in health plans.
- The measure is not risk adjusted and SDS factors were not assessed for use in the measure.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- This measure excludes hospitalizations in which the patient was transferred directly to a non-acute care facility for any diagnosis, patients with asthma (diagnosis or medications), COPD, obstructive chronic bronchitis, hypotension/heart block >1 degree or sinus bradycardia, and patients who are identified as having an intolerance or allergy to beta-blocker therapy.
- The developer states there are no exclusions in 2b3, which is inconsistent with information provided in S.10 and S.11
- The developer does not provide information on exclusions testing and analysis.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This intermediate clinical outcome measure is not risk adjusted.

Questions for the Committee:

- *Do you agree with the developer that risk adjustment is not necessary for this measure?*

2b5. Meaningful difference:

- The overall mean performance on this measure is 83% (SD = 7%) for the commercial health plans, 84% (SD=10%) for Medicaid and 90% (SD=6%) for Medicare for 2014.
- The difference between 25th and 75th percentile is statistically significant for all product lines (Commercial, Medicaid, and Medicare). The largest performance gap (11%) is in Medicaid plans. Overall, results suggest there are meaningful differences in performance and there is an opportunity for improvement.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- The developer states that this measure is collected with a complete sample, there is no missing data on this measure.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- Measure specifications are clearly defined and consistent with the evidence. Measure likely to be implemented consistently. Methodology for determining "persistence" explained, but no definition of "allowed gap days."
- Denominator: concerned if there is actually a way to determine all AMIs. What if the MI goes on for CABG or Stent. Is their diagnosis now Stent or MI? Will the denominator capture these AMIs? High reliability testing reported (.81)

2a2.: ***Committee's Comments on Reliability-Testing:***

- Reliability testing was done through beta-binomial model measuring signal-to-noise ratio, and the results demonstrated moderate reliability.
- Overall .78-.81 yes

2b1.: ***Committee's Comments on Validity-Specifications***:

- Measure specifications are clearly defined and consistent with the evidence. Measure likely to be implemented consistently. Methodology for determining "persistence" explained, but no definition of "allowed gap days."

- Expert panel determined this was valid, empirical testing demonstrated a moderate correlation with other measures including comprehensive diabetes care, cholesterol management and phamacotherapy with COPD exacerbation

2b2.: ***Committee's Comments on Validity-Testing:***

- Both construct and face validity done. Construct validity done through Pearson correlation testing and results indicate moderate and statistically significant correlation. Face validity also done through three expert panels and all unanimously supported measure. No statistical results of face validity testing provided. Overall moderate validity.
- Mostly expert panel determined this was valid

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Exclusions noted, though inconsistency in application. No apparent threats, but meaningful differences are noted in clinician performance. No risk adjustment done.
- According to the boxplots there is a 7-11% gap in performance between first quartile and third quartile. Largest gap is in Medicaid plans. there are meaningful differences, but small

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- The data source includes administrative claims, electronic clinical data, including pharmacy claims and readily available data occurring during patient care. All data elements are in defined fields in electronic claims.
- The developers do not provide specific information on the operational use of the measure; instead they outline the HEDIS Compliance Audit process to verify that HEDIS specifications are met. In addition to the audit, NCQA provides a system that allows for 'real-time' feedback from measure users.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- Data collection obtained through administrative claims, EHRs, and other sources. No data collection barriers identified. No assessments of costs or abstraction time done. Moderate to high feasibility.
- Data from electronic clinical data, pharmacy and administrative claims. Could be put to operational use, but caution regarding denominator and identifying individuals with AMI.

Criterion 4: Usability and Use

4. Usability and 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.

- This measure is publicly reported in NCQA's Quality Compass and Annual State of Health Care Quality. It is also

used in scoring for accreditation of Medicare Advantage Health Plans.
- The developer reports performance improvements over the past 3 years ~ 2% annually for Medicare, Medicaid and commercial plans, and no unintended consequences were identified from measure use.

Questions for the Committee:
- *Is the measure publicly reported?*
- *Is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments
Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***
- Measure is currently reported publically through NCQA Health Plan Rankings, Accreditation, and Quality Compass. High usability.
- Benefits outweigh unintended consequences at this time. Publicly reported on Quality Compass and Annual State of health care quality

Criterion 5: Related and Competing Measures

- 0070 : Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction
- The developer states that the differences between measures 0071 and 0070 do not have an impact on interpretability of publically reported rates, or the burden of data collection, because all data for both measures are collected from different data sources by different entities.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0071

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

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

Date of Submission: 6/30/2015

Instructions
- *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.*

- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*includes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☒ Intermediate clinical outcome (*e.g., lab value*): A 180-day course of treatment with beta-blockers

☐ Process:

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

N/A

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

N/A

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Patient 18 years of age and older is hospitalized>>> Health care provider diagnoses patient with acute myocardial infarction (AMI)>>> Health care provider and patient discuss the risk and benefits of beta-blocker therapy post discharge>>> Patient is dispensed a 180-day course of treatment with beta-blockers>>> Persistent beta-blocker use in patient's treatment reduces the risk of mortality, reduces the risk and severity of reinfarction, and improves the preservation of the left ventricular function>>> Improvement in quality of life and functioning for patient (Desired outcome).

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

ST-Elevation Myocardial Infarction (STEMI) Guideline:

O'Gara PT, Kushner FG, Ascheim DD, et al. 2013 ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction : A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2013;61(4):e78-e140. doi:10.1016/j.jacc.2012.11.019

- URL: http://content.onlinejacc.org/article.aspx?articleid=1486115

Non-ST Elevation Myocardial Infarction (NSTEMI) Guideline:

Amsterdam EA, Wenger NK, Brindis RG, et al. 2014 AHA/ACC Guideline for the Management of Patients With Non–ST-Elevation Acute Coronary Syndromes: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2014;64(24):2645-2687. doi:10.1016/j.jacc.2014.09.016.

- URL: http://content.onlinejacc.org/article.aspx?articleid=1910085&resultClick=3

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Guideline	Recommendation	Verbatim Quote	Estimate of Certainty (Precision) of Treatment	Size of Treatment Effect
STEMI Guideline	1	"Beta blockers should be continued during and after hospitalization for all patients with STEMI and with no contradictions to their use."	Level B	Class I
NSTEMI GUIDELINE	2	"In patients with concomitant NSTE-ACS [non-ST-elevation acute coronary syndrome], stabilized HF [heart failure], and reduced systolic function, it is recommended to continue beta blocker therapy with 1 of the 3 drugs proven to reduce mortality in patients with HF: sustained-release metoprolol succinate, carvedilol, or bosoprolol."	Level C	Class I
NSTEMI GUIDELINE	3	"It is reasonable to continue beta blocker therapy in patients with normal LV [left ventricular] function with NSTE-ACS"	Level C	Class lla

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

Both STEMI and NSTEMI guidelines use the same grading system.

Guideline	Recommendation	Size of Treatment	Definition of Class

		Effect (Grade)	
STEMI	1	Class I	Benefit > > > Risk Procedure/Treatment SHOULD be performed/administered
NSTEMI	2	Class I	Benefit > > > Risk Procedure/Treatment SHOULD be performed/administered
NSTEMI	3	Class IIa	Benefit > > Risk Additional studies with focused objective needed **IT IS REASONABLE** to perform procedure/administer treatment

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Size of Treatment Effect (Grade)	Definition of Class
Class IIb	Benefit ≥ Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment **MAY BE CONSIDERED**
Class III- No benefit	No Benefit Procedure/Test- Not Helpful Treatment- No proven benefit
Class III - Harm	Harm Procedure/Test – Excess cost without benefit or harmful Treatment- Harmful to patients

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Both guidelines cited in 1a.4.1 use the same methodology for grading recommendations.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Both guidelines used to support this measure cover a much wider topic area than just secondary prevention of myocardioal infarction with persistent beta-blocker therapy treatment and do not discuss in detail the evidence review process for each recommendation supporting the persistence of beta-blocker treatment after heart attack measure. They do, however, provide a grade of evidence for each of the recommendations and cite systematic reviews supporting those recommendations. Therefore, to answer the questions in 1a.7, we are using the evidence grades the guidelines provide and referencing one seminal systematic review cited in the guidelines that summarizes the body of evidence supporting the recommendations.

- Freemantle N, Cleland J, Young P, Mason J, Harrison J. Beta blockade after myocardial infarction: systematic review and meta regression analysis. BMJ. 1999;318:1730–1737.

http://www.ncbi.nlm.nih.gov/pubmed/10381708

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Both guidelines cited in 1a.4.1 use the same methodology for grading evidence.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

The evidence for this measure focuses on the importance of beta blocker therapy in long-term secondary prevention of acute myocardial infarction (AMI). It is important to note that a systematic evidence review completed by Freemantle et al., in 1999 supports and is referenced by both STEMI and NSTEMI guidelines.

Freemantle et al. assessed the effectiveness of beta-blockers in longer-term secondary prevention of AMI using randomized controlled trials (RCTs). The review focused on RCTs that compared beta-blockers to placebo.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Guideline	Recommendation	Estimate of Certainty (Precision) of Treatment	Definition of Level:
STEMI Guideline	**1**	**B**	Limited populations evaluated* Data derived from a single randomized trial or nonrandomized studies
NSTEMI Guideline	**2**	**C**	Very limited populations* Only consensus opinion of experts, case

			studies, or standard of care
NSTEMI Guideline	**3**	**C**	Very limited populations* Only consensus opinion of experts, case studies, or standard of care

*Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Estimate of Certainty (Percision) of Treatment	Definition of Level:
A	Multiple populations evaluated* Data derived from multiple randomized clinical trials or meta-analyses

*Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range:

It should be noted that the body of evidence supporting the guideline recommendations is much broader and includes more recent evidence than the evidence used in the Freemantle et al. systematic review which includes studies published from 1966-1997.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

There are 31 long-term randomized controlled trials included in the systematic evidence review by Freemantle et al. (1999), which supports the STEMI and NSTEMI guideline recommendations regarding persistent beta-blocker treatment after heart attack.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

The Freemantle et al. systematic evidence review rated the quality of studies as reasonably high, with adequate follow-up achieved in many trials.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/decline across studies, results of meta-analysis, and statistical significance*)

Considerable evidence supports the routine long term use of beta blockers in patients who have had a myocardial infarction, with substantial benefits in terms of reduced mortality and morbidity.

Freemantle et al., use a random effects approach in long term trials for incidence of risk difference to estimate to normalized annual reduction in mortality across trials. This approach suggests an annual reduction of 1.2 deaths in 100 patients treated with beta-blockers after myocardial infarction; that is about 84 patients will require treatment for 1 year to avoid one death. A similar approach was used to estimate the effects of treatment on reinfarction, although only 21 of the 34 comparisons provided data on reinfarction, resulting in wider confidence intervals and the potential for reporting bias. This analysis suggests an annual reduction in reinfarction of 0.9 events every 100 (0.3 to 1.6); that is about 107 patients would require treatment of 1 year to avoid one non-fatal reinfarction. There was a 23% reduction in the odds of death in long term trials (95% confidence interval 15% to 31%).

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

The guidelines and systematic review provide extremely limited findings regarding harm associated with persistent beta blocker treatment after a heart attack. Freemantle et al., studied withdrawal from treatment for both active treatment and placebo groups. The trials reported that dizziness, depression, cold extremitites, and fatigue were only marginally more common in the treatment than control groups. This supports the fact that the benefits of beta-blocker treatment significantly outweigh the minor treatment harms.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

There have been many (>100) studies examining the use of beta blockers in patients who have had an MI since the publication of the systematic reviews used to generate the STEMI and NSTEMI guidelines. An article published in 2012 by Bangalore et al., confirms that beta-blockers remain the standard of care after a myocardial infarction. This study also references the findings from the Freemantle et al., systematic review used to support the recommendations for our measure.

Bangalore S, Steg G, Deedwania P, et al. β-Blocker Use and Clinical Outcomes in Stable Outpatients With and Without Coronary Artery Disease. *JAMA.* 2012;308(13):1340-1349. doi:10.1001/jama.2012.12559.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
FINAL_Evidence_Form_0071_PBH.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
This measure addresses the appropriate clinical management of a person who has experienced an AMI. Persistent beta-blocker treatment after a heart attack reduces the risk of mortality, reduces the risk and severity of reinfarction, and improves the preservation of the left ventricular function.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The following data are extracted from HEDIS data collection and reflect the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by the mean, standard deviation, performance percentiles (10th, 25th, 50th, 75th and 90th percentile) and the interquartile range. Data is stratified by year and product line (i.e. commercial, Medicare, Medicaid, HMO and PPO) at the health plan level.

Persistence of Beta-Blocker Treatment After a Heart Attack

N=Number of Health Plans

Commercial (HMO and PPO Combined)
YEAR|N|MEAN|ST DEV|10TH| 25TH| 50TH|75TH|90TH|Interquartile Range
2012|250|79%|8%|68%|73%|79%|84%|89%|11%
2013|252|81%|8%|71%|77%|82%|87%|91%|10%
2014|253|83%|7%|73%|78%|82%|88%|91% |10%

Medicaid
YEAR|N|MEAN|ST DEV|10TH|25TH|50TH|75TH|90TH|Interquartile Range
2012|54|80%|10%|67%|73%|83%|88%|91%|15%
2013|64|82%|9%|71%|78%|83%|88%|91%|10%
2014|75|84%|10%|72%|80%|86%|91%|95%|11%

Medicare
YEAR|N|MEAN|ST DEV|10TH|25TH|50TH|75Th|90TH|Interquartile Range
2012|255|87%|6%|79%|84%|88%|91%|94%|7%
2013|269|89%|6%|81%|85%|90%|93%|95%|8%
2014|269|90%|6%|83%|87%|91%|94%|96%|7%

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

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
HEDIS data are stratified by type of insurance (e.g. commercial, Medicaid, Medicare). NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities. Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.

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

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Heart disease is the leading cause of death for people of most ethnicities in the United States, including African Americans, Hispanics, and whites. For American Indians or Alaska Natives and Asians or Pacific Islanders, heart disease is the second leading cause of death (CDC, 2015). Non-Hispanic black adults are at least 50% more likely to die of heart disease or stroke prematurely (i.e., before age 75 years) than their non-Hispanic white counterparts (CDC, 2013). Black women and men are more likely to die before age 75 as a result of coronary heart disease (CHD) than white women and men (rates of death are 37.9%, 61.5%, 19.4%, and 41.5%, respectively) (CDC, 2011). Racial and age-related disparities also exist in rates of recurrent MI or fatal CHD within 5 years of a first MI. Of those who have a first MI, the percentage with a recurrent event is as follows: at 45 to 64 years of age, 14% of white men, 18% of white women, 22% of black men, and 28% of black women; at =65 years of age, 21% of white men and women, 33% of black men, and 26% of black women (Mozaffarian et al., 2015).

A 2012 study by Zhang et al. compared medication adherence among MI survivors by disability, status, race/ethnicity, and income for all Medicare fee-for-service beneficiaries discharged post-MI in 2008. Among the disabled who were taking beta-blockers, the percentage of beneficiaries with good adherence for 6-month adherence was highest for Whites at 67% and lowest for Blacks at 52% with Asians, Hispanics, and Native Americans ranging in between. (Zhang et al., 2012). Moreover, 1 year adherence rates among beneficiaries were 68%, 66%, 61%, 58%, and 57% for Whites, Asians, Hispanics, Native Americans and Backs respectively. A comparison by income showed that among dual eligibles, Black beneficiaries were less likely to have good adherence compared to White beneficiaries at both 6 months and 1-year post-MI (Zhang et al., 2012).

The CDC analyzed data from 2008-2012 to identify if employment status had an impact on rates of CHD/stroke. The results of this analysis showed that 1.9% of employed adults aged <55 years reported a history of CHD/stroke, compared with 2.5% of unemployed adults looking for work, and 6.3% of adults not in the labor force. Workers employed in service and blue collar occupations were more likely than those in white collar occupations to report a history of CHD/stroke (Luckhaupt, 2014).

Center for Disease Control and Prevention (CDC). 2015. Heart Disease Facts. Last modified February 19, 2015. http://www.cdc.gov/heartdisease/facts.htm

Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services. 2013. "CDC Health Disparities and Inequalities Report-United States, 2013." Morbidity and Mortality Weekly Report (MMWR) 62(03); 1-2. http://www.cdc.gov/mmwr/pdf/other/su6203.pdf

Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services. 2011. "Fact Sheet: Health Disparities in Coronary Heart Disease and Stroke." http://www.cdc.gov/minorityhealth/CHDIR/2011/FactSheets/CHDStroke.pdf

Luckhaupt, S.E., Calvert, G.M., August 2014. "Prevalence of Coronary Heart Disease or Stroke Among Workers Aged <55 years- United States 2008-2012." Morbidity and Mortality Weekly Report (MMWR) 63(30); 645-649. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6330a1.htm

Mozaffarian, D., Benjamin E.J., Go A.S., et al. 2015. "Heart disease and stroke statistics—2015 update: a report from the American Heart Association." Circulation. 131:e29-e322. doi: 10.1161/CIR.0000000000000152

Zhang Y, Baik SH, Chang C-CH, Kaplan CM, Lave JR. Disability, Race/ethnicity, and Medication Adherence Among Medicare Myocardial Infarction Survivors. American heart journal. 2012;164(3):425-433.e4. doi:10.1016/j.ahj.2012.05.021.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality, High resource use

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Heart disease is the leading cause of death in the United States (CDC, 2015). Approximately every 34 seconds, someone in the United States has a heart attack, and approximately every 90 seconds, an American dies of one (Mozaffarian, et al., 2015). Every year an estimated 735,000 Americans have a heart attack. Of these, 210,000 have had at least one prior heart attack, while the remaining 525,000 have their first heart attack (Mozaffarian, et al., 2015). About 15% of people who have a heart attack will die from it (Mozaffarian, et al., 2015). High blood pressure, high cholesterol and smoking can lead to heart disease and heart attack. Nearly half of Americans adults have one or more of these risk factors (CDC, 2011).

The estimated annual costs for cardiovascular disease (CVD) and stroke in the United States for 2011 was $320.1 billion. This figure includes $195.6 billion for in direct costs and $124.5 billion in indirect costs from lost future productivity. CVD costs more than any other diagnostic group (Mozaffarian, et al., 2015).

Beta-blocker treatment decreases the likelihood of sudden cardiac death (Kezerashivili, 2012).
Beta-blockers are heart medications that work, in part, by lowering the heart rate, which reduces the amount of force on the heart and blood vessels (AHA, 2013). Treatment with beta-blockers after a heart attack can help reduce blood pressure (AHA, 2014). Around 70% of people who have a heart attack also have high blood pressure (Mozaffarian, et al., 2015). Although beta blockers can reduce chest pain and decrease the likelihood of a future heart attack, only 52% of heart-attack patients adhere to their beta-blocker treatment at 180 days post discharge. (AHA, 2015. Kramer, et al 2006).

1c.4. Citations for data demonstrating high priority provided in 1a.3

American Heart Association (AHA). 2015. Cardiac Medications. Last modified May 6, 2015. http://www.heart.org/HEARTORG/Conditions/HeartAttack/PreventionTreatmentofHeartAttack/Cardiac-Medications_UCM_303937_Article.jsp

American Heart Association (AHA). 2013. "How do beta blocker drugs affect exercise?" last modified June 9, 2014. http://www.heart.org/HEARTORG/Conditions/More/MyHeartandStrokeNews/How-do-beta-blocker-drugs-affect-exercise_UCM_450771_Article.jsp

Centers for Disease Control and Prevention (CDC). 2015. "Leading Causes of Death." Last modified February 6, 2015. http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

Centers for Disease Control and Prevention (CDC). 2011. "Million Hearts: Strategies to reduce the prevalence of leading cardiovascular disease risk factors–United States, 2011." Morbidity and Mortality Weekly Report (MMWR) 60(36):1248–51

Kezerashivili, A., K. Marzo, J. De Leon. 2012. Beta Blocker Use After Acute Myocardial Infarction in the Patient with Normal Systolic Function: When is it "OK" to Discontinue?. Current Cardiology Reviews; 8(1):77-84. www.ncbi.nlm.nih.gov/pmc/articles/PMC3394111/pdf/CCR-8-77.pdf

Kramer, J,M., B. Hammill, K.J. Anstrom, et al. September 2006. National evaluation of adherence to B-blocker therapy for 1 year after acute myocardial infarction in patients with commercial health insurance. America Heart Journal; (454): e1-8. DOI: 10.1016/j,ahj.2006.02.030

Mozaffarian, D., Benjamin E.J., Go A.S., et al. 2015. "Heart disease and stroke statistics—2015 update: a report from the American Heart Association." Circulation. 131:e29-e322. doi: 10.1161/CIR.0000000000000152

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

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 subcriteria 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 : Acute Myocardial Infarction, Prevention

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention

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.)*
http://www.ncqa.org/tabid/59/Default.aspx

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** 0071_PBH_Value_Sets_Final.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date

and explain the reasons.
There are no significant changes to the measure specification since the last endorsement maintenance completed on January 18, 2012.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who had a 180-day course of treatment with beta-blockers post discharge.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator: 12 month measurement period (calendar year) plus a 6 month look back period.
Denominator: A 12 month period beginning on July 1 of the year prior to the measurement year through June 30 of the measurement year.

S.6. **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, 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.
ADMINISTRATIVE
Patients who had a 180-day course of treatment with beta-blockers post-discharge. Post discharge refers to patients discharged from an acute inpatient setting with an AMI (AMI Value Set) from July 1 of the year prior to the measurement year through June 30 of the measurement year.

In order to identify patients with "persistent" beta-blocker treatment, identify all patients in the denominator population whose dispensed days supply is =135 days in the 180-day measurement interval. The measure defines persistence of treatment as at least 75 percent of the days supply filled.

To determine continuity of treatment during the 180-day period, identify all prescriptions filled within the 180-day measurement interval, and add the number of allowed gap days (up to 45 days) to the number of treatment days for a maximum of 180 days (i.e., 135 treatment days + 45 gap days = 180 days).

To account for patients who are on beta-blockers prior to admission, factor those prescriptions into adherence rates if the actual treatment days fall within the 180-day measurement interval.

DEFINITIONS
Treatment days (days covered) - The actual number of calendar days covered with prescriptions within the specified 180-day measurement interval (i.e., a prescription of a 90-day supply dispensed on the 100th day will have 80 days counted in the 180-day interval).

180-day measurement interval - The 180 day period that includes the discharge date and the 179 days after discharge.

TABLE PBH-B BETA-BLOCKER MEDICATIONS
DESCRIPTION / PRESCRIPTION
Noncardioselective beta-blockers / Carvedilol; Labetalol; Nadolol; Penbutolol; Pindolol; Propranolol; Timlol; Sotalol
Cardioselective beta-blockers / Acebutolol; Atenolol; Betaxolol; Bisoprolol; Metoprolol; Nebivolol
Antihypertensive combinations / Atenolol-chlorthalidone; Bendroflumethiazide-nadolol; Bisoprolol-hydrochlorothiazide; Hydrochlorothiazide-metoprolol; Hydrochlorothiazide-propranolol

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients 18 years of age and older as of December 31 of the measurement year who were hospitalized and discharged from July 1 of the year prior to the measurement year to June 30 of the measurement year with diagnosis of AMI. See question S.9 Denominator Details for methods to identify patients who qualify for the denominator.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Patients discharged from an acute inpatient setting with an AMI (AMI Value Set) from July 1 of the year prior to the measurement year through June 30 of the measurement year.

Use only facility claims to identify denominator events (including readmissions or direct transfers). Do not use professional claims.

If a patient has more than one episode of AMI from July 1 of the year prior to the measurement year through June 30 of the measurement year, only include the first discharge.

Transfers to acute facilities: Include hospitalizations in which the patient was transferred directly to another acute inpatient facility for any diagnosis. Count the discharge from the subsequent acute inpatient facility, not the initial discharge. The discharge date from the facility to which the patient was transferred must occur on or before
June 30 of the measurement year.

Transfers to nonacute facilities. Exclude from the denominator, hospitalizations in which the patient was transferred directly to a nonacute care facility for any diagnosis.

Readmissions: If the patient was readmitted to an acute or nonacute care facility for any diagnosis, include the patient in the denominator and use the discharge date from the original hospitalization.

Due to the extensive volume of codes associated with identifying the denominator for this measure, we are attaching a separate file with code value sets. See code value sets located in question S.2b.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclude from the denominator, hospitalizations in which the patient was transferred directly to a nonacute care facility for any diagnosis.

Exclude patients who are identified as having an intolerance or allergy to beta-blocker therapy. Any of the following anytime during the patient's history through the end of the continuous enrollment period meet criteria:
- Asthma (Asthma Value Set).
- COPD (COPD Value Set).
- Obstructive chronic bronchitis (Obstructive Chronic Bronchitis Value Set).
- Chronic respiratory conditions due to fumes and vapors (Chronic Respiratory Conditions Due to Fumes/Vapors Value Set).
- Hypotension, heart block >1 degree or sinus bradycardia (Beta-Blocker Contraindications Value Set).
- A medication dispensing event indicative of a history of asthma (Table PBH-D).
- Intolerance or allergy to beta-blocker therapy.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
MEDICATIONS TO IDENTIFY EXCLUSIONS (History of Asthma)
DESCRIPTION / PRESCRIPTION
Bronchodilator combinations / Albuterol-ipratropium; Budesonide-formoterol; Fluticasone-salmeterol; Mometasone-formoterol
Inhaled corticosteroids / Beclomethasone; Budesonide; Ciclesonide; Flunisolide; Fluticasone; Fluticasone CFC free; Mometasone; Triamcinolone

Due to the extensive volume of codes associated with identifying denominator exclusions for this measure we are attaching a separate file with code value sets (except for medications to identify patients with a history of asthma). See code value sets located in question S.2b.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*

N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
STEP 1. Determine the eligible population. To do so, identify patients who meet all specified criteria.
-AGES: 18 years and older as of December 31 of the measurement year.
-EVENT/DIAGNOSIS: Identify patients who were discharged from an acute setting with an AMI (AMI Value Set) from July 1 of the year prior to the measurement year through June 30 of the measurement year. Use only facility claims.
STEP 2: Exclude patients who meet the exclusions criteria. SEE S.10 AND S.11 FOR DENOMINATOR EXCLUSION CRITERIA AND DETAILS.
STEP 3: Determine the number of patients in the eligible population who were given a 180-day course of treatment with beta blockers post discharge.
STEP 4: Identify patients whose dispensed days supply is = 135 days in the 180-day measurement interval
STEP 5: Calculate the rate by dividing the numerator (Step 4) by the denominator (after exclusions) (Step 2).

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.

N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Pharmacy

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Organizations via NCQA's online data submission system.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Integrated Delivery System

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
FINAL_Testing_Form_0071_PBH.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0071
Measure Title: Persistence of Beta-Blocker Treatment After a Heart Attack
Date of Submission: 6/30/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs

☒ other: Electronic Clinical Data: Pharmacy	☒ other: Electronic Clinical Data: Pharmacy

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
N/A
1.3. What are the dates of the data used in testing? Click here to enter date range

Health Plan Level (2014) HEDIS
Testing of performance measure score with beta binomial reliability was performed with data from measurement year 2013.

Testing of construct validity with Pearson's Correlation was performed using data from measurement year 2013.

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.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☒ other: Integrated Delivery System	☒ other: Integrated Delivery System

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*)
Measure Score Reliability Testing and Construct Validity Testing
Measure score reliability and construct validity were calculated from U.S. HEDIS data that included all health plans submitting data to NCQA for HEDIS: 253 Commercial plans, 75 Medicaid plans and 269 Medicare plans. The plans were geographically diverse and varied in size.

Systematic Assessment for Face Validity
This measure was tested for face validity with three expert panels. See additional information: Ad.1. Workgroup/Expert Panel in Measure Development for names and affiliations of expert panels:

1. Cardiovascular Measurement Advisory Panel includes eight physicians and one nurse with expertise in cardiovascular health and quality measurement.
2. The Technical Measurement Advisory Panel includes 12 members, including representation by health plans, methodologists, clinician and auditors.
3. NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that

performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

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

Patient Samples for Measure Score Reliability and Construct Validity Testing

Health Plan Level

In measurement year 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Data are summarized at the health plan level and stratified by product line (i.e. commercial, Medicaid, Medicare). Below is a description of the sample. It includes the number of health plans included in HEDIS data collection and the average eligible population for the measure across health plans.

Product Line	Number of Plans	Average Denominator Size per Plan
Commercial (HMO/PPO combined)	253	157
Medicaid	75	83
Medicare	269	171

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

Reliability testing of the measure score and construct validity of this measure were tested using the same sample. Validity was also demonstrated through a systematic assessment of face validity. The composition of each panel is described in question 1.5 above.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Data are stratified by commercial, Medicaid and Medicare product lines.

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*)

Performance Measure Score (Beta Binomial)

In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009). The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

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

Results of Reliability Testing of Performance Measure Score (Beta Binomial):

Health Plan Level (Measurement Year 2013)

Below, we present health plan level data, which includes data submitted from 253 commercial plans, 75 Medicaid plans, and 269 Medicare plans.

Commercial		Medicaid		Medicare	
Overall	10th-90th	Overall	10th-90th	Overall	10th-90th
0.78	0.42-0.89	0.81	0.66-0.92	0.78	0.42-0.91

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?*)

Performance Measure Score (Beta Binomial):

Reliability scores can vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (signal). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities.

Health Plan Level

Testing suggests that this measure has good reliability at the health plan level between 0.78 and 0.81. The 10th-90th percentile distribution of health plan level reliability on this measure shows there is large variation. This large variation is due to a lower number of plans reporting the rate overall because of small denominators for the plans. It is important to note that overall health plans have met or exceeded the minimally accepted threshold of 0.7 and many of the plans exceeded 0.8. Good reliability is demonstrated since variations in these large populations are due to signal and not to noise.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Empirical Validity:

Method of testing construct validity: At the health plan level, we tested for construct validity by exploring whether Persistence of Beta-Blocker Treatment after a Heart Attack (#0071) correlated with the following measures:

- Comprehensive Diabetes Care (CDC) LDL-C: The percentage of adults 18-75 years of age with diabetes who had LDL-C control <100 mg/dL during the measurement year.
- Cholesterol Management for Patients with Cardiovascular Conditions (CMC): The percentage of adults 18-75 years of age discharged with an AMI, coronary artery bypass graft (CABG) or percutaneous coronary interventions (PCI) in the year prior to measurement year, or who had a diagnosis of ischemic vascular disease (IVD) during the measurement year and the year prior to the measurement year, and who had LDL-C control (<100 mg/dL).
- Pharmacotherapy Management of COPD Exacerbation (PCE): The percentage of COPD exacerbations for adults 40 years of age and older who had an acute inpatient discharge or ED visit on or between January 1-November 30 of the measure year and who were dispensed appropriate medications.

To test these correlations we used a Person correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

Systematic Assessment of Face Validity:

Health Plan Level

Method of Assessing Face Validity: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. Refer to What Makes a Measure "Desirable"? The work-up is vetted by NCQA's Measurement Advisory Panels (MAPs), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures.
NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQAs Board of Directors will be included in the next HEDIS year and reported as first-year measures.

STEP 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. The first-year distinction guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was already tested as part of its development—rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

STEP 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be publicly reported and may be used for scoring in accreditation.

Step 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, NCQA prioritizes measures for re-evaluation and selected measures are researched for changes in clinical guidelines or in the health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.

Expert Participation
This measure was tested for face validity with input from two expert panels. Updated guidelines from the American Heart Association and American College of Cardiology in 2013 and 2014 were also a strong authoritative source in applying the evidence for our persistence of beta-blocker treatment after a heart attack measure

We list an overview of each panel here. Please refer to Ad.1 in the submission form for the names and affiliation of experts in each panel.

1.) Cardiovascular Measurement Advisory Panel includes eight physicians and one nurse with expertise in cardiovascular health and quality measurement.
2.) The Technical Measurement Advisory Panel includes 12 members, including representation by health plans, methodologists, clinician and auditors.
3.) NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

ICD-10 CONVERSION:

Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

Steps in ICD-9 to ICD-10 Conversion Process

1. NCQA staff identify ICD-10 codes to be considered based on ICD-9 codes currently in measure. Use GEM to identify ICD-10 codes that map to ICD-9 codes. Review GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA staff identify additional codes (not identified by GEM mapping step) that should be considered. Using ICD-10 tabular list and ICD-10 Index, search by diagnosis or procedure name for appropriate codes.
3. NCQA HEDIS Expert Coding Panel review NCQA staff recommendations and provide feedback.
4. As needed, NCQA Measurement Advisory Panels perform clinical review. Due to increased specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is intended to be included in the scope of the measure. Not all ICD-10 recommendations are reviewed by NCQA MAP; MAP review items are identified during staff conversion or by HEDIS Expert Coding Panel.
5. Post ICD-10 code recommendations for public review and comment.
6. Reconcile public comments. Obtain additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
7. NCQA staff finalize ICD-10 code recommendations.

Tools Used to Identify/Map to ICD-10

All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).
GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Empirical Validity

Health Plan Level

Correlations between Persistence of Beta-Blocker Treatment after a Heart Attack (PBH) and Comprehensive Diabetes Control (CDC) LDL-C, Cholesterol Management for Patients With Cardiovascular Conditions (CMC), and Pharmacotherapy Management of COPD Exacerbation (PCE) measures (2013)

	Person Correlation Coefficients

Quality Measure	CDC-LDL-C	CMC	PCE-Systemic	PCE-Bronchodi	PBH
PBH (Commercial)	0.38094	0.30215	0.25865	0.33448	1
PBH (Medicaid)	0.4031	0.32383	0.47421	0.51638	1
PBH (Medicare)	0.41992	0.36396	0.47394	0.42068	1

Note: All correlations are significant at p<.05

Systematic Assessment of Face Validity

Expert Panels

Our expert panels unanimously supported the necessity of persistent beta-blocker treatment after a heart attack to measure quality of care in patients with cardiovascular disease.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Empirical Validity

Health Plan

Correlations testing suggests that the PBH measure has a moderate correlation with the diabetes and cardiovascular disease cholesterol control measures and COPD medication measure, as hypothesized. This positive correlation indicates that plans that are performing well on measures of evidence-based care, are doing so across their members with cardiovascular (PBH, CMC), diabetic (CDC LDL-C), and COPD (PCE) conditions. In other words, plans that perform well on the PBH measure also perform well on the cardiovascular disease (CMC) and diabetes (CDC LDL-C) cholesterol control measures, which focus on cholesterol management for patients with cardiovascular disease and diabetes. Similarly, those same plans perform well on COPD medication measure (PCE) which assesses the provision of appropriate medications for COPD patients who have an exacerbation. The p value of all correlation analyses are <.05 which indicates that the correlations are statistically significant.

Systematic Assessment of Face Validity

Expert Panels

Multiple NCQA panels concluded with good agreement that the measure as specified accurately assesses persistence of beta blocker use for at-risk patients in health plans. Our interpretation of these results is that this measure has sufficient face validity.

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

2b3.2. 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)

2b3.3. 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*)

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than .05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the 75th percentile of performance. We used these two plans as examples of measured entities. However, the method can be used for comparison of any two measured entities.

2b5.2. 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*)

Health Plan Level: HEDIS 2014 Variation in Performance Across Health Plans

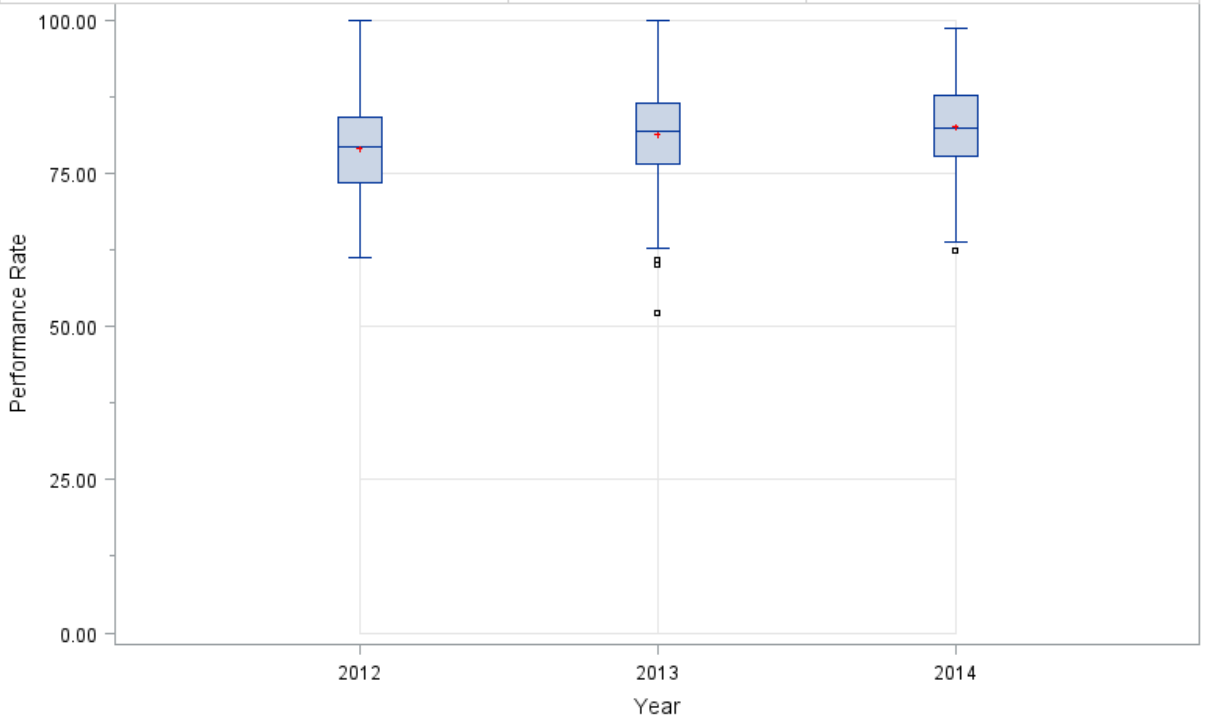
	N	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial	253	83%	7%	73%	78%	82%	88%	91%	10%	<0.05
Medicaid	75	84%	10%	72%	80%	86%	91%	95%	11%	<0.05
Medicare	269	90%	6%	83%	87%	91%	94%	96%	7%	<0.05

N= Number of plans reporting
IQR: Interquartile range
p-value: p-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

Box plots for HEDIS 2014 Variation in Performance Across Health Plans are included below for your reference.

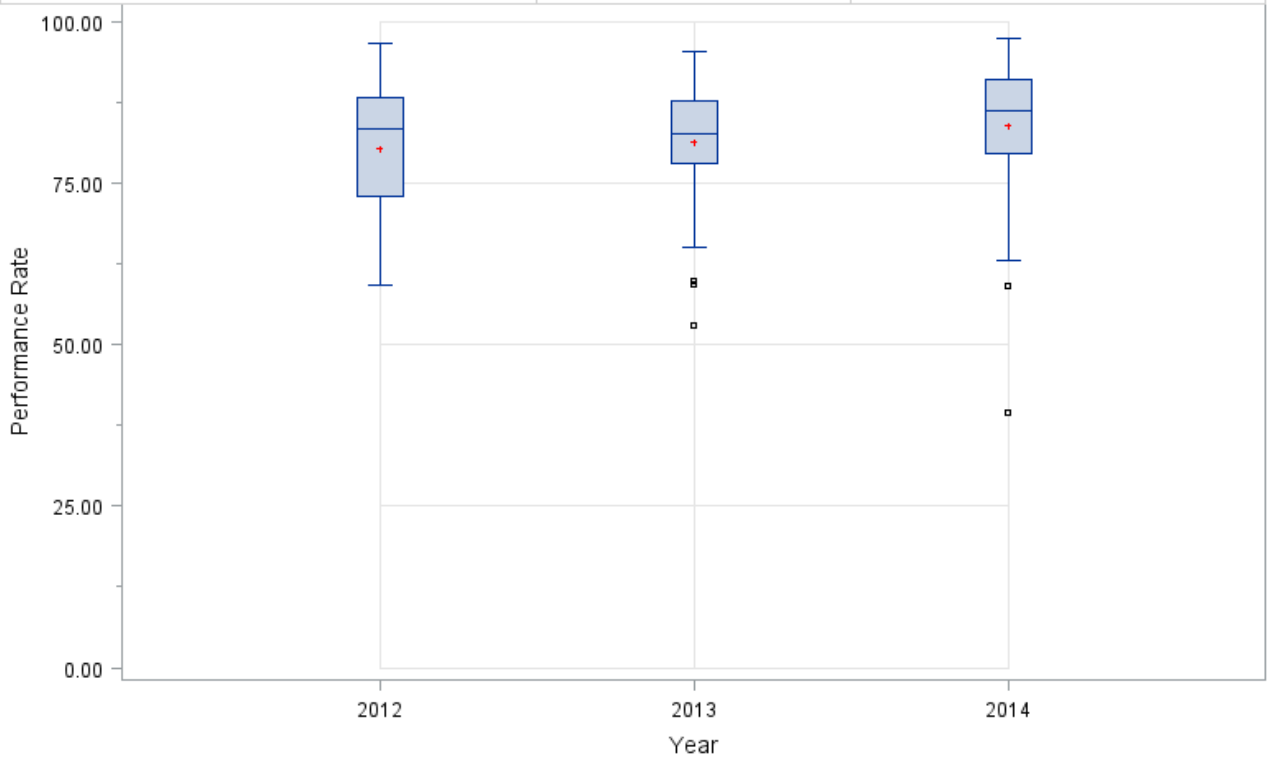
Boxplot Graph for Commercial Persistence of Beta-Blocker Treatment After a Heart Attack from HEDIS 2012-2014

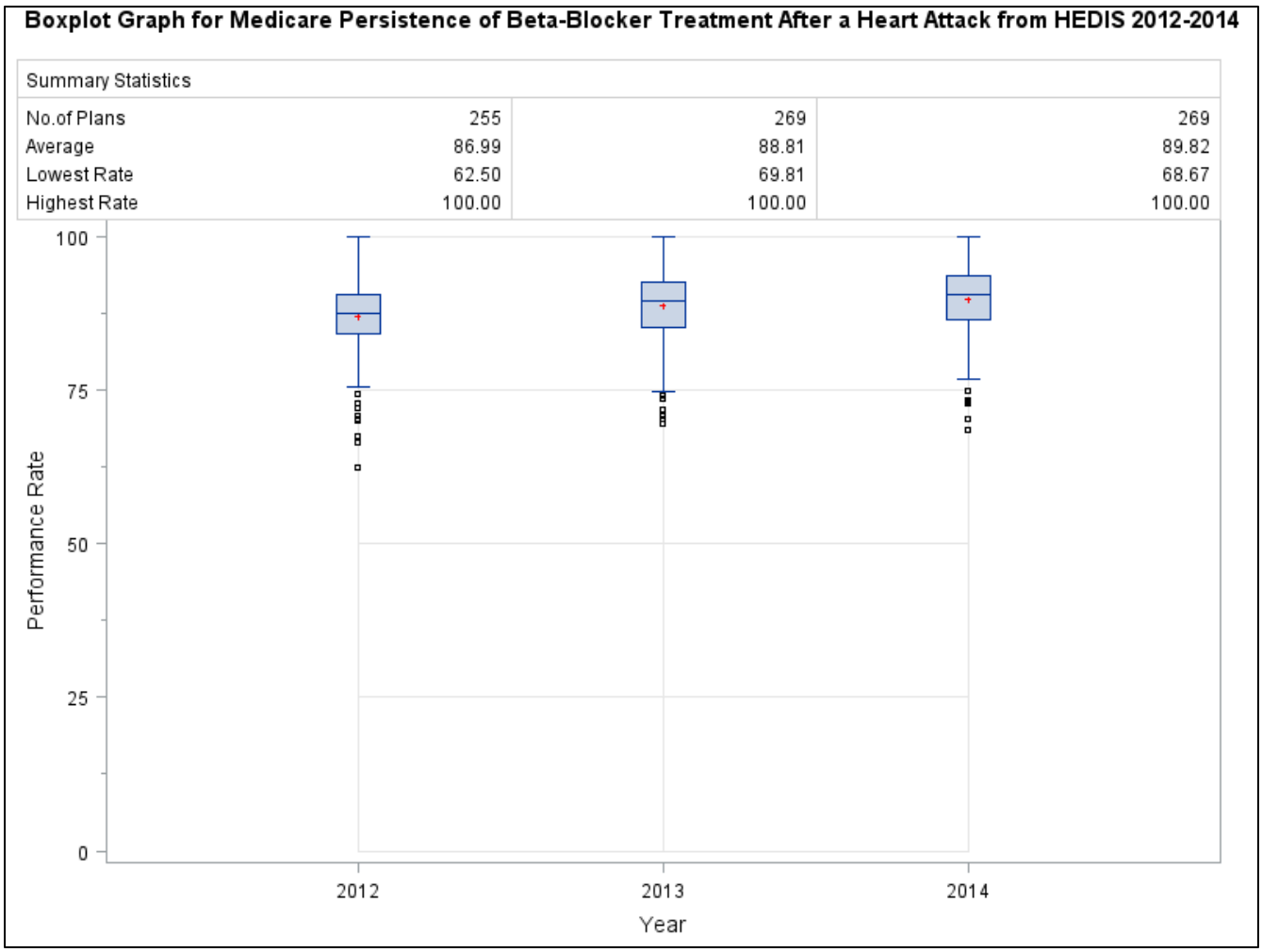
Summary Statistics			
No.of Plans	250	252	253
Average	79.06	81.49	82.62
Lowest Rate	61.29	52.27	62.50
Highest Rate	100.00	100.00	98.68



Boxplot Graph for Medicaid Persistence of Beta-Blocker Treatment After a Heart Attack from HEDIS 2012-2014

Summary Statistics			
No.of Plans	54	64	75
Average	80.49	81.50	83.88
Lowest Rate	59.15	53.14	39.53
Highest Rate	96.74	95.45	97.56





2b5.3. 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?*)

Health Plan Level

The results indicate there is a 7-11% gap in performance between 25th and 75th performing plans. The largest performance gap (11%) is in Medicaid plans. The difference between 25th and 75th percentile is statistically significant for all product lines (Commercial, Medicaid, and Medicare). There is also a 13-23% gap in performance between the 10th and 90th performing plans. Overall, results suggest there are meaningful differences in performance and there is an opportunity for improvement.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
N/A

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)
This measure is collected with a complete sample, there is no missing data on this measure.

2b7.2. 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*)
This measure is collected with a complete sample, there is no missing data on this measure.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)
This measure is collected with a complete sample, there is no missing data on this measure.

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)
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)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
NCQA recognizes that, despite the clear specifications defined for HEDIS measures, data collection and calculation methods may vary, and other errors may taint the results, diminishing the usefulness of HEDIS data for managed care organization (MCO) comparison. In order for HEDIS to reach its full potential, NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.
The HEDIS Compliance Audit addresses the following functions:
1) Information practices and control procedures
2) Sampling methods and procedures
3) Data integrity
4) Compliance with HEDIS specifications
5) Analytic file production
6) Reporting and documentation
In addition to the HEDIS Audit, NCQA provides a system to allow “real-time” feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system is vital to the regular re-evaluation of NCQA measures.
Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
Broad public use and dissemination of this measure is encouraged. NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care providers in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

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.

Planned	Current Use (for current use provide URL)
	Public Reporting Health Plan Ranking; Annual State of Health Care Quality Health Plan Ranking; Annual State of Health Care Quality http://reportcard.ncqa.org/plan/external/plansearch.aspx http://www.ncqa.org/ReportCards/HealthPlans/StateofHealthCareQuality.aspx Regulatory and Accreditation Programs NCQA Accreditation http://www.ncqa.org/tabid/123/Default.aspx Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Compass; Annual State of Health Care Quality http://www.ncqa.org/tabid/177/Default.aspx Quality Compass; Annual State of Health Care Quality http://www.ncqa.org/ReportCards/HealthPlans/StateofHealthCareQuality.aspx

4a.1. **For each CURRENT use, checked above, provide:**

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

HEALTH PLAN RANKING: NCQA ranks health plans using the methodology we have used every year since 2005. For the 2014-2015 rankings, NCQA studied almost 1,400 health plans and ranked more than 1,000 of them based on clinical performance, member satisfaction and results from NCQA Accreditation surveys

ANNUAL STATE OF HEALTH CARE QUALITY: NCQA produces the State of Health Care Quality Report yearly to focus on key quality issues facing the United States and to drive improvement in the delivery of evidence-based medicine. The report documents performance trends over time, tracks variations in care and recommends quality improvements. The 2014 report provides data for the 2013 calendar year. Data in the report comes from 814 HMOs and 353 PPOs, representing more than 171 million people or 54 percent of the U.S. population.

HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2013, a total of 170 Medicare Advantage health plans were accredited using this measure among others covering 7.1 million Medicare beneficiaries. This measure is also used in scoring for accreditation of 336 commercial health plans covering 87 million lives; 77 Medicaid health plans covering 9.1 million lives. Health plans are scored based on performance compared to benchmarks.

QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program,**

certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

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

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Performance across all plan types has steadily improved over the past three years, with Medicare, Medicaid, and commercial plan performance increasing each year by about 2%. Within each percentile (10th, 25th, 50th etc.), there is a steady performance increase over the past three years and it is important to note even the lowest performers have a steady increase in performance. Current average performance (2014) is highest in Medicare plans (90%), followed by Medicaid plans (84%) and then commercial plans (83%).

4b.2. **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.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences to individuals identified during the testing and long-standing use of this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0070 : Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)

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

5a. Harmonization

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 completely harmonized?
No

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.
DUE TO THE TEXT LIMIT IN THIS SECTION – WE ARE PROVIDING OUR ANSWER FOR 5a.2 IN SECTION 5b.1

5b. Competing Measures

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

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
ANSWER FOR SECTION 5a.2

NCQA's current Persistence of Beta Blocker Treatment After a Heart Attack measure (NQF measure 0071) uses health plan-reported data to assess the percentage of patients 18 years of age and older during the measurement year who were discharged with a diagnosis of AMI during the 6 months prior to the beginning of the measurement year through the 6 months after the beginning of the measurement year and who received persistent beta-blocker treatment for six months after discharge.

RELATED NQF MEASURE 0070:
This measure assesses the percentage of patients aged 18 years and older with a diagnosis of coronary artery disease seen within a 12 month period who also have a prior MI or a current left ventricular ejection fraction (LVEF) <40% who were prescribed beta-blocker therapy.

HARMONIZED MEASURE ELEMENTS:
Measure 0071 and 0070 focus on patients 18 years and older who are prescribed beta-blocker treatment post-discharge after having a MI or history of MI. The National Quality Strategy Priorities classification for both measures is Prevention and Treatment of Cardiovascular Disease. Both measures exclude patients who are allergic or have an intolerance to beta blockers.

UNHARMONIZED MEASURE ELEMENTS:
Below are the unharmonized measure elements between measure 0071 and measure 0070:

Measure 0071 focuses on beta-blocker treatment post a MI and Measure 0070 focuses on patients who have a prior MI or a current or prior LVEF <40%.

-Data Source: Data for measure 0071 is collected through administrative claims, electronic clinical data, and pharmacy data, while data for measure 0070 is collected through medical record, electronic health record data, electronic clinical data, and paper records

-Level of Accountability: Measure 0071 is a health plan level measure while measure 0070 is a clinician-level measure.

-Population: Measure 0071 focuses on patients who were diagnosed with a MI and discharged and prescribed a beta-blocker therapy treatment. Measure 0070 focuses on patients in a measurement year with a diagnosis of coronary artery diseases who also have a prior MI or current or prior LVEF

-Exclusions: The difference in exclusions is that measure 0071 specifies asthma, COPD, obstructive chronic bronchitis, chronic

respiratory conditions due to fumes and vapors, hypotension, heart block >1 degree, sinus bradycardia, and medication dispensing events indicative of a history of asthma as exclusions. Additionally, measure 0071 excludes hospitalizations in which the patient was transferred directly to a nonacute care facility for any diagnosis. Measure 0070 exclusions include: documentation of patient reason(s) for not prescribing beta-blocker therapy (e.g., patient declined, other patient reasons) and documentation of system reason(s) for not prescribing beta-blocker therapy (e.g., other reasons attributable to the health care system

IMPACT ON INTERPRETABILITY AND DATA COLLECTION BURDEN:
The differences between measures 0071 and 0070 do not have an impact on interpretability of publically reported rates, or the burden of data collection, because all data for both measures are collected from different data sources by different entities.

ANSWER FOR SECTION 5b.1

Our current measure has a long standing history of use by health plans and has been implemented for 10 years.

Appendix

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

Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** National Committee for Quality Assurance
Co.2 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-
Co.3 **Measure Developer if different from Measure Steward:** National Committee for Quality Assurance
Co.4 **Point of Contact:** Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
NCQA follows a standard process of vetting members of the measurement advisory panel for conflicts of interest.
Measure Developer/Steward Updates and Ongoing Maintenance

CARDIOVASCULAR MEASUREMENT ADVISORY PANEL
Stephen Persell, MD, MPH (Chair), Northwestern University
Tom Kottke, MD, HealthPartners
Eduardo Ortiz, MD, MPH, Wolters Kluwer Health
David Goff, Jr., MD, PhD, FAHA, FACP, University of Colorado
Kathy Berra, MSN, ANP, FAAN, Stanford University
Michael Pignone, MD, MPH, University of North Carolina at Chapel Hill
Randall S. Stafford, MD, PhD, Stanford University
Tracy Wolff, MD, Agency for Healthcare Research and Quality
Corinne Husten, MD, MPH, U.S. Food and Drug Administration

TECHNICAL MEASUREMENT ADVISORY PANEL
Andy Amster, MSPH, Kaiser Permanente
Kathryn Coltin, MPH, Harvard Pilgrim Health Care
Lekisha Daniel-Robinson, MSPH, Centers for Medicare & Medicaid Services
Marissa M. Finn, MBA, Cigna HealthCare
Scott Fox, MS, Med, Independence Blue Cross
Carlos Hernandez, CenCal Health
Kelly Isom, RN, MA, Aetna
Harmon S. Jordan, ScD, RTI International

Ernest Moy, MD, MPH, Agency for Healthcare Research and Quality (AHRQ)
Patrick Roohan, New York State Department of Health (NYSDOH)
Lynne Rothney-Kozlak, MPH, Rothney-Kozlak Consulting, LLC
Natan Szapiro, Independence Blue Cross

COMMITTEE ON PERFORMANCE MEASUREMENT
Peter Bach, MD, Memorial Sloan Kettering Cancer Center
Bruce Bagley, MD, TransforMED
Andrew Baskin, MD, Aetna
A. John Blair lll, MD, Taconic IPA, Inc
Patrick Conway, MD, MSC, Center for Medicare & Medicaid Services
Jonathan D. Darer, MD, Geisinger Health System
Helen Darling, National Business Group on Health
Foster Gesten, MD, NYSDOH Office of Managed Care
Marge Ginsburg, Center for Healthcare Decisions
Christine Hunter, MD, (Co-Chair) US Office of Personnel Management
Jeffrey Kelman, MMSc, MD, Centers for Medicare & Medicaid Services
Philip Madvig, MD, The Permanente Medical Group
J. Brent Pawlecki, MD MMM, The Goodyear Tire & Rubber Company
Susan Reinhard, RN, PhD, AARP
Eric C. Schneider, MD, MSc (Co-Chair), RAND Corporation
Marcus Thygeson, MD, MPH Blue Shield of California

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2005

Ad.3 **Month and Year of most recent revision:** 07, 2009

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

Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Ad.8 **Additional Information/Comments:** Publication of each Measure is to be accompanied by the following notice:

NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.
These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA.

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 #: 0079
Measure Title: Heart Failure: Left Ventricular Ejection Fraction Assessment (Outpatient Setting)
Measure Steward: American College of Cardiology
Brief Description of Measure: Percentage of patients aged 18 years and older with a diagnosis of heart failure for whom the quantitative or qualitative results of a recent or prior (any time in the past) LVEF assessment is documented within a 12 month period.
Developer Rationale: This measure is aimed at improving the number of patients with heart failure who receive an evaluation of their LVEF.

Measurement of LVEF in heart failure patients is key to the implementation of therapeutic interventions demonstrated to slow disease progression and improve outcomes in patients with left ventricular systolic dysfunction.

Numerator Statement: Patients for whom the quantitative or qualitative results of a recent or prior (any time in the past) LVEF assessment is documented* within a 12 month period.

*Documentation must include documentation in a progress note of the results of an LVEF assessment, regardless of when the evaluation of ejection fraction was performed.

Qualitative results correspond to numeric equivalents as follows:
Hyperdynamic: corresponds to LVEF greater than 70%
Normal: corresponds to LVEF 50% to 70% (midpoint 60%)
Mild dysfunction: corresponds to LVEF 40% to 49% (midpoint 45%)
Moderate dysfunction: corresponds to LVEF 30% to 39% (midpoint 35%)
Severe dysfunction: corresponds to LVEF less than 30%
Denominator Statement: All patients aged 18 years and older with a diagnosis of heart failure.
Denominator Exclusions: None.

Measure Type: Process
Data Source: Electronic Clinical Data : Registry
Level of Analysis: Clinician : Individual
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: 8/10/09 **Most Recent Endorsement Date:** 1/18/12

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

Public and Member Comment

- Although this measure is intended for an outpatient setting, in the numerator it states that documentation must include documentation in a progress note of the results of an LVEF assessment, regardless of when the evaluation of ejection fraction was performed, which may involve documentation of an LVEF from an in-patient hospital setting. In-patient hospital data may not be readily available.
- It is a waste of resources to collect and report on mere completion of an assessment.
- Request clarification in the specifications about EFs done in prior visits or documented in the Electronic Health Record. A

provider may acknowledge these procedures, but not provide billing codes for a visit done in the office/outpatient setting.

- Functional outcomes such as this are the primary correlate of health-related quality of life (HRQL). HRQL is now recognized as the key patient-centered outcome. Thus, to measure only the indicators of provider care without acknowledging the patients perspective seems ill-advised. I strongly encourage you to reconsider this stance.

Developer response:

- While the measure requires that a patient's LVEF status be documented at least once within a 12 month period, the measure does not specify a time period for the assessment of LVEF - this assessment may have been performed anytime previously or within the last 12 months. Evaluation of LVEF in patients with heart failure provides important information that is required by any clinician managing the patient's outpatient care to appropriately direct treatment.
- This measure is intended to encourage assessment of a patient's LVEF status in order to identify patients who may be candidates for particular therapeutic options. An EHR could be searched for the relevant data to determine results of a previous LVEF assessment. For claims-based reporting, a provider would have to document the results of an LVEF assessment, regardless of when the evaluation of ejection fraction was performed.
- This is an assessment measure, not an outcome measure. The assessment only, without regard to subsequent intervention or follow-up is not proximal to the outcome which is the actual functional status of the patient.

Steering Committee: Reviewed comments and developers responses. No change in recommendations.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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.

The developer provides the following evidence for this process measure:

- This is a clinician level process measure that calculates the % patients aged 18 years and older with a diagnosis of heart failure for whom the quantitative or qualitative results of a recent or prior (any time in the past) LVEF assessment is documented within a 12 month period in a clinician office/clinic setting.
- The evidence recommends initial evaluation and repeat measurement with a change in patient status.
- The developer provides decision logic from assessment to outcome for the evaluation of LVEF to slowed disease progression and improved health outcomes.
- The developer provides one guideline with two guideline statements for the management of heart failure with the details of the Quantity/Quality/Consistency for the 2013 ACCF/AHA guideline. The guideline statements are both graded **Class I: Level of Evidence: C** or consensus opinion of experts, case studies, or standard of care.
- The developers state that the two articles published after the publication of the 2013 ACCF/AHA guideline both further support the assessment of EF and guideline-directed medical treatment.

Questions for the Committee:

- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*

- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
- *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- The developer provided 2013 and 2014 performance data from the PINNACLE registry:

	Mean	Std Dev	Min	Max	Interquartile score	# providers	# patients	Male	White	Black	Other
2013	67.8%	32.0%	0.00%	100%	49.6%	2254	409,332	55.8%	89.5%	8.7%	1.9%
2014	72.5%	29.9%	0.00%	100%	41.6%	2219	404,406	55.7%	90.4%	7.2%	2.4%

- A 2003 study of patients within 1 month of treatment initiation indicates 35.2% of patients with CHF received an evaluation of LVEF. Another study cited found that between 2002 and 2003, 84% of inpatients (rather than the outpatient population for this measure) in hospitals had documentation of LVEF assessment.
- Variation in provider performance based on sex, age, race, and insurance is provided. The sample populations also list body mass index (BMI), diabetes, hypertension, Atrial Fibrillation/Flutter, Heart Failure, Peripheral Arterial Disease, Stroke/TIA and Myocardial Infarction co-morbidities apparently for information purpose.
- Data collected by AHRQ from 2001 to 2010, demonstrates disparities among Blacks and lower income residents with CHF despite overall improvement in the quality of care in patients with this diagnosis.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - For Measure 0079 (outpatient assessment of LVEF in heart failure patients), there are listed studies and consensus documents indicating the importance of this process measure but there are several important caveats:
     1. This is a process measure intended to support evidence-based clinical decision making. There is not a measure for the decision-making element of heart failure.
     2. The measure data can only reliably come from PINNACLE registry sources, which is limited in its scope so primary care physicians and a large percentage of cardiologists do not use this registry. Those physicians could not have valid data obtained.
     3. The source behind this measure is largely consensus driven with only one cited meta-analysis. As a process measure that is consensus developed it is weaker than an evidence-based outcome measure.
   - On the positive side, the ACC has been using this within the PINNACLE framework. So the greater the adoption of PINNACLE the greater the value of this measure.
   - This is a process measure. The measure is only supported by expert opinion in the heart failure guideline; however it is unlikely that higher level evidence will be available in the future. It is widely accepted that the patient's LVEF in particular is a critical value to consider when determining therapeutic options for patient with HF.
   - The two articles offered as evidence are not direct evidence of the utility of EF determination in HF patients.

One article is about the additional information CMR provides over and above echo. The other deals with predicting mortality of post STEMI patients based on EF -- not specifically HF patients.
- There are more proximal processes (eg. prescription of ACEIs or ARBs) but the EF value is important to determining the appropriateness of prescribing. However reporting the EF and doing something based on the EF are two different things.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***
- The developers cite on meta-analysis and several observational studies for this process measure. The use of left ventricular ejection fraction to guide therapy is included in ACC/AHA guidelines. However I am not aware that the frequency of testing has been defined especially for those with lower grades of heart failure who are clinically stable.
- The body of evidence provided is related to ACEI/ARB use not about performing or recording EF values. No detail is provided other than the types of studies. There was 1 meta-analysis

1b. ***Committee's Comments on Performance Gap:***
- The developers do provide substantial data over two separate time periods involving >4K "clinicians" (presumably only cardiologists) and nearly a half million patients. The developers do clearly show a gap in this performance. What is not clear is whether the appropriate clinical decision making occurred in the absence of documentation of the measurement.
- The cited study from 2003 is irrelevant. Much has changed regarding the management of HF in the past 12 years.
- The Pinnacle Registry information from 2013 and 2014 both show a significant performance gap (67.8% and 72.5%). It should be noted that the Pinnacle Registry patients were 89.5%-90.4% White.
- Data regarding disparities was cited from AHRQ but pertains to hospital admissions for HF, not the indicator being evaluated.
- There were disparities based on insurance provider in the 2013 data with Medicare patients having significantly lower % of EF evaluations recorded (none -69.1%; private-64.4%; Medicaid 60%; Medicare 47.6%)"

1c. ***Committee's Comments on Composite Performance Measure:***
- Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications
2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. • The measure's data source is the PINNACLE Registry with the data dictionary and collection tool provided. HCPCS codes provided for the numerator; ICD-9, ICD-10 and CPT codes provided for the denominator. ICD-10 conversion methodology and calculation algorithm are not provided. • For the numerator, documentation must include documentation in a progress note of the results of an LVEF assessment, regardless of when the evaluation fraction was performed. Theoretically, a provider could continue to successfully report the measure in subsequent measurement periods without additional assessment or interventions, if LVEF was assessed anytime during the patient's history. • LVEF is determined by either quantitative evaluation or narrative findings. The measure includes both concepts. ***Questions for the Committee:*** ○ *Are all the data elements clearly defined? Are all appropriate codes included?* ○ *Is the logic or calculation algorithm clear?* ○ *Is it likely this measure can be consistently implemented?*
2a2. Reliability Testing Testing attachment

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.

- Reliability testing was conducted on the performance score, using PINNACLE data source at the individual clinician level during calendar year 2013 for 2,254 providers and 409,332 patients; & 2014 for 2,219 providers and 404,406 patients.
- A signal-to-noise analysis using the beta-binomial model was conducted. This type of analysis, which is appropriate for the measure, quantifies the amount of variation in performance that is due to differences between providers (as opposed to differences that are due to random measurement error). 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. The method results in a reliability statistic for each clinician.
- As the minimum number of patient visits required (>10) the average reliability was 0.988 for 2013 & 0.989 for 2014. For providers with the median number of patient encounters, the reliability was even higher, with 0.997 for both years. A reliability of 0.70 is generally considered a minimum threshold for acceptability, and 0.80 is considered very good reliability. Very high quartiles reliability statistics are also provided for both 2013 and 2014.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The clinical practice guidelines supporting this measure recommend measurement of LVEF in patients with CHF during initial evaluation and periodically as symptoms or a change in status warrants.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- Content validity was assessed by an expert work group, public comment, concurrent formal peer review process, approval by the ACC Board of Trustee & Science Advisory & Coordination Committees, and PCPI membership.
- The developers state that construct validity (the degree to which the measure is assessing what it claims to be measuring) was difficult to assess as independent auditing has not occurred, though as the data is abstracted from the EHR by direct transfer, errors would occur by mapping or incorrect auditor abstracting.
- Face validity was systematically assessed via survey by 2 committees (one ACC & one AHA) of 42 members with 85.7% agreeing the measure scores as specified provide an accurate reflection of quality and can be used to distinguish good and poor quality. Using a Likert Scale from 1-5, the mean importance rate was 4.24.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- There are no exclusions for this measure.

2b4. Risk adjustment:

- This process measure is not risk adjusted.

2b5. Meaningful difference:

- The overall mean performance on this measure is 67.8% (SD = 32.0%) for 2013 and 72.5% (SD = 29.9%) for 2014.
- The developer notes a moderate to large amount of variability across providers for statistically 'identical' patients and suggests that a patient presenting to 1 provider, as opposed to another, would on average, be 2.3 times (2013) and 1.2 times (2014) more likely to have their EF assessed.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- As the measure is specified for Registry use, comparability across data sources/methods is not applicable.

2b7. Missing Data

- The developer notes that in the PINNACLE registry, if a data field is not completed, it is assumed the process of care was not done; therefore, missing numerator data is reported as a quality failure.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- The developers have done a very thoughtful job in developing the specifications so that reliability can be ensured--but that is only for users of PINNACLE. Other cardiologists and primary care physicians would have to manually extract the coding from clinical records with an anticipated reduction in reliability. In other words, it is the structure of the registry that ensures this to be a reliable measure. Outside of that particular registry the reliability may fall secondary to human and charting frailties.
- "Documentation in the progress note"" is somewhat vague. Examples of types of qualitative and quantitative entries are provided for the numerator. I am not sure how Pinnacle is able to extract this information but apparently it can. Possibly this is because the EHR in use is a cardiology application.
- The measure seems to be consistently implemented using Pinnacle and compatible EHRs.

2a2.: ***Committee's Comments on Reliability-Testing:***

- Yes, the reliability for the cardiologists participating with PINNACLE was more than adequate.
- There were over 2,000 providers and 400,000 patients included in each of the reliability testing periods -- adequate. Providers were nationwide. One downside - very low percentage of Black patients included in analyses.
- The signal-to-noise analysis was 0.988 for 2013 and 0.989 for 2014 (very good). A minimum of more than 10 patients was required to establish this level of reliability. Given the incidence of HF, this shouldn't be a problem. Most practices will have more than 10 HF patients.

2b1.: ***Committee's Comments on Validity-Specifications***:

- Since this measure is largely consensus derived process measure, the validity becomes dependent upon the coding associated with the consensus and not that based upon clinical evidence. As such there may be face validity but not statistical validity.
- The HF guidelines recommend determining the EF initially and with any significant change in symptoms or change in condition. Level of evidence is expert opinion.
- The specifications of the measure only require that the EF vlaue be mentioned in the documentation. No timing is specified other than 1 instance of documentation in 12 months. This is sufficiently vague to allow for EF determinations older than 12 months to be credited. However, there are no stipulations about timing in relationship to changes in patient condition. This is a weakness of this measure.

2b2.: ***Committee's Comments on Validity-Testing:***

- As above, the validity was determined through consensus and the testing of the data was performed with that bias included.

- See above. Content validity - expert workgroup; peer review
- Face validity - 1 AHA committee and 1 ACC committee agreed that measure socres reflect quality (85.7%) Importance ranking 4.24/5
- I see the recording of EF as being tangential to actions that could indicate quality (eg. prescription of ACEI/ARBs to patients with EF <40%)"

2b3-7.: ***Committee's Comments on Threats to Validity:***

- This measure was developed within the PINNACLE registry and was thoughtfully developed for that registry. Its appropriate application as a measure of process quality therefore is dependent upon the greater adoption of the registry as a tool. As such the greatest threat to the measure is its symbiotic relationship to PINNACLE.
- This is also critical to the intention of this process measure as a piece of information important to the clinical decision-making of the clinician in appropriately treating the patient with heart failure.
- 2b3 - no exclusions
- 2b4 - might need risk adjustment in populations other than the Pinnacle Registry as socioeconomic factors might impact patients' ability to access the test (echo). No risk adjustment provided with application
- 2b5 - there is enough variability among providers (according to developers) to distinguish differences in quality
- 2b6 - intended to be used with Pinnacle Registry
- 2b7 - missing data is interpreted as not fulfilling the criteria (poorer quality)

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- The data source is electronic abstraction from the clinical record, MDS, OASIS to the PINNACLE Registry from readily available data occurring during patient care.
- As one of the earliest NQF endorsed measures utilized, the developers state the data collection strategy is implemented, though no information for abstraction time and costs are provided.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- This measure is feasible and has been used--but only for those physicians using the PINNACLE registry. It is not feasible for those who do not use PINNACLE. Success of this process measure should be determined by either greater adoption of PINNACLE, incorporation of this into other EHRs, or development of extraction of the data from submitted medical claims.
- No feasibility data is provided.
- EF values are generally determined. Whether they are referenced in the Progress Notes on a routine basis or in a consistent way is unknown.
- May require an abstractor if Pinnacle and compatible EHRs are not used.

Criterion 4: Usability and Use

4. Usability and 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.

- The measure is currently used in the ACC PINNACLE Registry for quality improvement.
- The developers state they continuously seek opportunities for use of this measure in public reporting and accountability programs, and do not have policies that would restrict use.
- NQF policy states that performance results are used in at least 1 accountability application within 3 years after initial endorsement and are publicly reported within 6 years after initial endorsement. This measure was initially endorsed in 2009.
- No unintended consequences have been identified, though the developer continuously monitors.
- In 2012, the Measure Applications Partnership (MAP) Clinician Workgroup supported the measure for the Meaningful Use (EHR Incentive Program) - Eligible Professionals. No comments were provided. In 2014, the Clinician Workgroup did not the measure for Physician Compare and Value-Based Payment Modifier Program stating the measure does not adequately address any current needs of the program.

Questions for the Committee *:*

- *Is the measure publicly reported?*
- *For maintenance measures – is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- This measure is not currently being widely used in public reporting. It is used by ACC for feedback and cardiologist self-improvement
- Not publically reported
- Not in an accountability application
- Performance could assist with promoting high quality, efficient care but there are measure that are closer to outcomes of interest (readmissions in 30 days; ACEI/ARBs prescribed, etc.)"

Criterion 5: Related and Competing Measures

- 0135 : Evaluation of Left ventricular systolic function (LVS)
- The developer states that the measure specifications are not completely harmonized because 0135 is inpatient based and focuses on the assessment occurring prior to discharge. 0079 looks at whether the assessment was performed during a 12 month period for a patient with a diagnosis of heart failure.

Pre-meeting public and member comments

-

Measure Number (*if previously endorsed*)**:** 0079

Measure Title: Heart Failure: Left Ventricular Ejection Fraction (LVEF) Assessment (Outpatient Setting)

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: 6/23/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across

Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Assessment of left ventricular ejection fraction

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

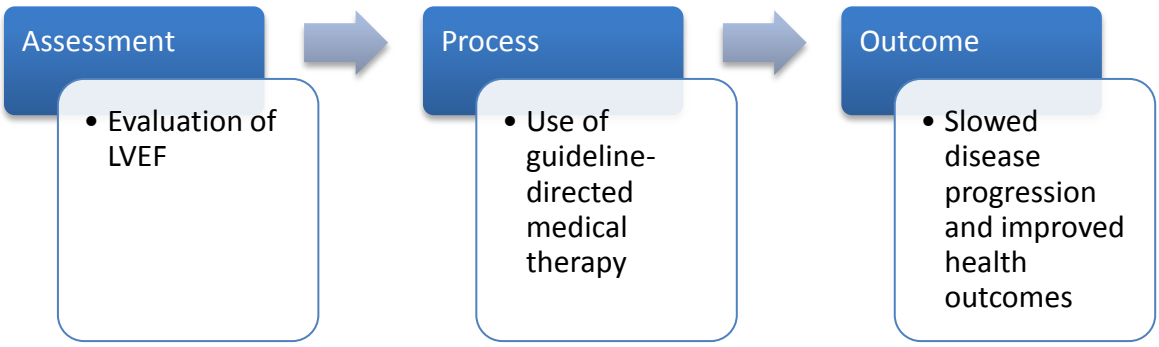
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



Evaluation of LVEF in patients with heart failure provides important information that is required to appropriately direct treatment. Several pharmacologic therapies have demonstrated efficacy in slowing disease progression and improving outcomes in patients with left ventricular systolic dysfunction. LVEF assessed

during the initial evaluation of patients presenting with heart failure can be considered valid unless the patient has demonstrated a major change in clinical status, experienced or recovered from a clinical event, or received therapy that might have a significant effect on cardiac function.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

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.

http://content.onlinejacc.org/article.aspx?articleid=1695825

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

2013 ACCF/AHA guideline for the management of heart failure (p. e165-166)

1. A 2-dimensional echocardiogram with Doppler should be performed during initial evaluation of patients presenting with HF to assess ventricular function, size, wall thickness, wall motion, and valve function. **Class I: Level of Evidence: C**
2. Repeat measurement of EF and measurement of the severity of structural remodeling are useful to provide information in patients with HF who have had a significant change in clinical status; who have experienced or recovered from a clinical event; or who have received treatment, including GDMT, that might have had a significant effect on cardiac function; or who may be candidates for device therapy. **Class I: Level of Evidence: C**

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Guideline Statement # (see 1a.4.2 above)	Class of Recommendation/Level of Evidence (for definitions see 1a.4.4 below)
1	Class Ic
2	Class Ic

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Class of Recommendation (COR) is an estimate of the size of the treatment effect considering risks versus benefits in addition to evidence and/or agreement that a given treatment or procedure is or is not useful/effective or in some situations may cause harm.

Class I: Procedure/Treatment should be performed/administered

Class IIa: It is reasonable to perform procedure/administer treatment

Class IIb: Procedure/Treatment may be considered

Class III: No benefit (Not helpful or No proven benefit)

Class III: Harm (Excess cost w/o benefit or Harmful to patients)

Specific COR definitions are included in Table 1 below.

Table 1. Applying Classification of Recommendation and Level of Evidence

SIZE OF TREATMENT EFFECT

ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT	CLASS I *Benefit >>> Risk* Procedure/Treatment **SHOULD** be performed/ administered	CLASS IIa *Benefit >> Risk* *Additional studies with focused objectives needed* **IT IS REASONABLE** to perform procedure/administer treatment	CLASS IIb *Benefit ≥ Risk* *Additional studies with broad objectives needed; additional registry data would be helpful* Procedure/Treatment **MAY BE CONSIDERED**	CLASS III *No Benefit* or CLASS III *Harm* Procedure/ Test / Treatment COR III: No benefit — Not Helpful — No Proven Benefit COR III: Harm — Excess Cost w/o Benefit or Harmful — 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
Suggested phrases for writing recommendations	should is recommended is indicated is useful/effective/beneficial	is reasonable can be useful/effective/beneficial is probably recommended or indicated	may/might be considered may/might be reasonable usefulness/effectiveness is unknown/unclear/uncertain or not well established	COR III: No Benefit: is not recommended; is not indicated; should not be performed/ administered/ other; is not useful/ beneficial/ effective COR III: Harm: potentially harmful; causes harm; associated with excess morbidity/mortality; should not be performed/ administered/ other
Comparative effectiveness phrases†	treatment/strategy A is recommended/indicated in preference to treatment B treatment A should be chosen over treatment B	treatment/strategy A is probably recommended/indicated in preference to treatment B it is reasonable to choose treatment A over treatment B		

Note: A recommendation with Level of Evidence B or C does not imply that the recommendation is weak. Many important clinical questions addressed in the guidelines do not lend themselves to clinical trials. Although randomized trials are unavailable, there may be a very clear clinical consensus that a particular test or therapy is useful or effective. *Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use. †For comparative effectiveness recommendations (Class I and IIa; Level of Evidence A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

ACCF/AHA Task Force on Practice Guidelines. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines. American College of Cardiology Foundation and American Heart Association, Inc. Cardiosource.com. 2010. Available at: http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf and http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

2013 ACCF/AHA Guideline for the Management of Heart Failure

This guideline covers multiple evaluation and management issues for the adult patient with Heart Failure (HF) including the assessment of left ventricular ejection fraction to enable appropriate treatment.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

An overall grade for the quality of evidence was not assigned. Rather, the quality of a study (or set of studies) supporting a recommendation was graded on an estimate of the certainty or precision of the treatment effect (see 1a.4.3).

Recommendations used to support this measure have a Level of Evidence C: Consensus opinion of experts, case studies, or standard of care.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Level of Evidence (LOE) is an estimate of the certainty or precision of the treatment effect.

Level of Evidence A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence C: Consensus opinion of experts, case studies, or standard of care.

Specific LOE definitions are included in Table 1 in 1a.4.4.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: An extensive evidence review was conducted through October 2011 and includes selected other references through April 2013.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

The body of evidence supporting the recommendations on ACE/ARB therapy includes:

4 observational

1 meta-analysis

1 prospective multicenter cohort

1 retrospective cohort

1 review paper

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

The recommendations for this process are rated as Level of Evidence C, meaning consensus opinion of experts, case studies, or standard of care. Additional information on the overall quality of evidence is not provided.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

2013 ACCF/AHA Guideline for the Management of Heart Failure (p. e165)

Although a complete history and physical examination are important first steps, the most useful diagnostic test in the evaluation of patients with or at risk for HF (e.g., postacute MI) is a comprehensive 2-dimensional echocardiogram; coupled with Doppler flow studies, the transthoracic echocardiogram can identify abnormalities of myocardium, heart valves, and pericardium. Echocardiography can reveal subclinical HF and predict risk of subsequent events. Use of echocardiograms in patients with suspected HF improves disease identification and provision of appropriate medical care.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

No harms have been identified related to the evaluation of patients with or at risk for HF. Rather, the 2013 ACCF/AHA Guideline for the Management of Heart Failure discusses whether certain testing modalities are superior to others with echocardiography being preferred "due to its widespread availability and lack of ionizing radiation" (p. e165).

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

One prospective study and one retrospective analysis were published after the publication of the 2013 ACCF/AHA Guideline for the Management of Heart Failure.

Note: Text below for description and results is verbatim from the article abstract.

Abbasi SA, Ertel A, Shah RV, Dandekar V, Chung J, Bhat G, Desai AA, Kwong RY, Farzaneh-Far A. Impact of cardiovascular magnetic resonance on management and clinical decision-making in heart failure patients. J Cardiovasc Magn Reson. 2013;15:89.

Description: We prospectively studied 150 consecutive patients with heart failure and an ejection fraction ≤ 50% referred for CMR [Cardiovascular magnetic resonance]. Definitions for "significant clinical impact" of CMR were pre-defined and collected directly from medical records and/or from patients. Categories of significant clinical impact included: new diagnosis, medication change, hospital admission/discharge, as well as performance or avoidance of invasive procedures (angiography, revascularization, device therapy or biopsy).

Results: Overall, CMR had a significant clinical impact in 65% of patients. This included an entirely new diagnosis in 30% of cases and a change in management in 52%. CMR results directly led to angiography in 9% and to the performance of percutaneous coronary intervention in 7%. In a multivariable model that included clinical and imaging parameters, presence of late gadolinium enhancement (LGE) was the only independent predictor of "significant clinical impact" (OR 6.72, 95% CI 2.56-17.60, p=0.0001).

Conclusion: CMR made a significant additive clinical impact on management, decision-making and diagnosis in 65% of heart failure patients. This additive impact was seen despite universal use of prior echocardiography in this patient group. The presence of LGE was the best independent predictor of significant clinical impact following CMR.

Liosis S, Bauer T, Schiele R, Gohlke H, Gottwik M, Katus H, Sabin G, Zahn R, Schneider S, Rauch B, Senges J, Zeymer U. Predictors of 1-year mortality in patients with contemporary guideline-adherent therapy after acute myocardial infarction: results from the OMEGA study. Clin Res Cardiol. 2013;102:671-7.

Description and Results: We performed a retrospective analysis of 3,782 patients surviving acute ST-elevation and non ST-elevation myocardial infarction who were enrolled in the prospective, randomized, double-blind, controlled OMEGA trial with 104 German centers. The primary objective of the OMEGA study was to determine the effect of highly purified omega-3 fatty acid ethyl esters-90 on the rate of sudden cardiac death in patients surviving AMI and receiving current guideline-adherent treatment within the 1-year of follow-up. 80.8 % of the patients received early revascularization therapy. At discharge, 94.2 % of the patients received beta-blocker, 90.4 % ACE inhibitor/angiotensin receptor blocker, 94.3 % statin, 95.4 % aspirin and 88.4 % clopidogrel. During the 1-year follow-up 139 patients (3.7 %) died. Multivariate logistic regression analysis revealed the following independent predictors of 1-year mortality in decreasing order of importance: ejection fraction <45 % [odds ratio (OR) 2.28, 95 % confidence interval (CI) 1.53-3.41], age ≥70 years (OR 2.17, 95 % CI 1.42-3.32), no acute revascularization (OR 2.02, 95 % CI 1.33-3.08), prior stroke/transient ischemic attack (OR 1.90, 95 % CI 1.09-3.30), peripheral arterial disease (OR 1.86, 95 % CI 1.12-3.10), heart rate >85/min (OR 1.82, 95 % CI 1.23-2.71), chronic obstructive lung disease (OR 1.77, 95 % CI 1.01-3.10) and HDL cholesterol <40 mg/dl (OR 1.75, 95 % CI 1.15-2.67).

Conclusion: In patients surviving AMI and treated with contemporary guideline-adherent therapy, 1-year mortality was low. Nevertheless, traditional risk factors such as ejection fraction <45 %, older age, no acute revascularization and comorbidities were the strongest predictors of long-term mortality supporting the findings from previous studies.

Impact on conclusions of systematic review: Both articles further support the assessment of ejection fraction to enable appropriate guideline-directed medical treatment.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
HF_LVEF_Assessment_0079_Evidence_Form_5_26_15.pdf

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., the benefits or improvements in quality envisioned by use of this measure)*
This measure is aimed at improving the number of patients with heart failure who receive an evaluation of their LVEF.

Measurement of LVEF in heart failure patients is key to the implementation of therapeutic interventions demonstrated to slow disease progression and improve outcomes in patients with left ventricular systolic dysfunction.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
2013 performance data from the Pinnacle registry.

Overall mean performance on this measure is 67.8%, with a standard deviation of 32.0%. The minimum score equals 0.00%, while the maximum score equals 100.00%. The interquartile score is equal to 49.6%.

2,254 providers were measured, and the patient study sample equals 409,332. 55.8% of the sample is male. 89.5% of the sample is white, 8.7% is black, and 1.9% identified as “other.” The sample reached across all US regions, with 14.9% of providers in the

Northeast, 30.7% of providers in the Midwest, 37.4% of providers in the South, and 17.1% of providers in the West.

Mean
Decile 1 4.0%
Decile 2 23.5%
Decile 3 44.5%
Decile 4 62.3%
Decile 5 75.5%
Decile 6 84.9%
Decile 7 90.6%
Decile 8 94.7%
Decile 9 97.7%
Decile 10 99.7%

2014 performance data from the Pinnacle registry.

Overall mean performance on this measure is 72.5%, with a standard deviation of 29.9%. The minimum score equals 0.00%, while the maximum score equals 100.00%. The interquartile score is equal to 41.6%.

2,219 providers were measured, and the patient study sample equals 404,406. 55.7% of the sample is male. 90.4% of the sample is white, 7.2% is black, and 2.4% identified as "other." The sample reached across all US regions, with 15.1% of providers in the Northeast, 24.5% of providers in the Midwest, 42.0% of providers in the South, and 18.5 % of providers in the West.

Mean
Decile 1 7.2%
Decile 2 34.7%
Decile 3 54.4%
Decile 4 69.9%
Decile 5 82.1%
Decile 6 89.4%
Decile 7 93.1%
Decile 8 95.9%
Decile 9 98.0%
Decile 10 99.7%

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 2003 study analyzing the quality of health care in the U.S. found that only 35.25% of participants with congestive heart failure who were beginning medical treatment received an evaluation of their LVEF within 1 month of the start of treatment.(1) For patients hospitalized with heart failure, a study analyzing data from 223 hospitals participating in the Acute Decompensated Heart Failure National Registry (ADHERE) between July 2002 and December 2003 found that left ventricular function assessment was documented in 84% of the 69,069 eligible admissions. Variability among participating hospitals was significant with rates at individual hospitals varying from 14 to 100%.(2)

(1)Appendix to McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. N Engl J Med. 2003;348:2635-2645.

(2)Fonarow GC, Yancy CW, Heywood JT. Adherence to heart failure quality-of-care indicators in US hospitals: analysis of the ADHERE Registry. Arch Intern Med. 2005; 165: 1469–1477.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
We examined variation in provider performance on this measure based on sex, age, race and a number of other patient factors to

identify variations. The findings fare represented for 2013 and 2014 respectively.

To view the tables in formatted fashion, see testing form Section 2b5.1 page 14.

2013

2013 stratified descriptive statistics of performance rate from Pinnacle Registry.

Male
# of providers: 2250
# of patients: 228280
Minimum: 0.00%
Lower Quartile: 47.6%
Mean: 69.3%
Upper Quartile: 95.5%
Maximum: 100%
Quartile Range: 47.8%
Std Dev: 31.8%

Female
# of providers: 2250
# of patients: 180550
Minimum: 0.00%
Lower Quartile: 41.0%
Mean: 66.2%
Upper Quartile: 94.4%
Maximum: 100%
Quartile Range: 53.4%
Std Dev: 32.8%

Age: <60
# of providers: 2236
# of patients: 77884
Minimum: 0.00%
Lower Quartile: 42.0%
Mean: 66.1%
Upper Quartile: 94.4%
Maximum: 100%
Quartile Range: 52.4%
Std Dev: 33.3%

Age: 60 -< 70
# of providers: 2246
# of patients: 96129
Minimum: 0.00%
Lower Quartile: 46.3%
Mean: 68.4%
Upper Quartile: 96.4%
Maximum: 100%
Quartile Range: 50.1%
Std Dev: 33.0%

Age: 70 -< 80
# of providers: 2253
# of patients: 120506
Minimum: 0.00%

Lower Quartile: 46.8%
Mean: 69.4%
Upper Quartile: 96.9%
Maximum: 100%
Quartile Range: 50.1%
Std Dev: 32.8%

Age: >= 80
# of providers: 2243
# of patients: 114813
Minimum: 0.00%
Lower Quartile: 42.9%
Mean: 67.3%
Upper Quartile: 96.0%
Maximum: 100%
Quartile Range: 53.1%
Std Dev: 33.4%

Insurance: None
# of providers: 69
# of patients: 311
Minimum: 0.00%
Lower Quartile: 25.0%
Mean: 69.1%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 75.0%
Std Dev: 42.7%

Insurance: Private
# of providers: 596
# of patients: 33938
Minimum: 0.00%
Lower Quartile: 33.3%
Mean: 64.4%
Upper Quartile: 97.1%
Maximum: 100%
Quartile Range: 63.8%
Std Dev: 35.7%

Insurance: Medicaid
# of providers: 382
# of patients: 13596
Minimum: 0.00%
Lower Quartile: 28.6%
Mean: 60.0%
Upper Quartile: 90.2%
Maximum: 100%
Quartile Range: 61.6%
Std Dev: 35.6%

Insurance: Medicare
# of providers: 23
# of patients: 107
Minimum: 0.00%
Lower Quartile: 0.00%
Mean: 47.6%

Upper Quartile: 91.7%
Maximum: 100%
Quartile Range: 91.7%
Std Dev: 44.2%

Insurance: Other
# of providers: 76
# of patients: 204
Minimum: 0.00%
Lower Quartile: 50.0%
Mean: 76.3%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 50.0%
Std Dev: 36.3%

Race: White
# of providers: 1479
# of patients: 202768
Minimum: 0.00%
Lower Quartile: 50.8%
Mean: 71.2%
Upper Quartile: 96.9%
Maximum: 100%
Quartile Range: 46.1%
Std Dev: 31.4%

Race: Black
# of providers: 1267
# of patients: 19665
Minimum: 0.00%
Lower Quartile: 48.6%
Mean: 70.5%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 51.4%
Std Dev: 36.3%

Race: Other
# of providers: 842
# of patients: 4211
Minimum: 0.00%
Lower Quartile: 33.3%
Mean: 67.2%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 66.7%
Std Dev: 40.6%

2014

2014 stratified descriptive statistics of performance rate from Pinnacle Registry.

Male
# of providers: 2214
# of patients: 224353

Minimum: 0.00%
Lower Quartile: 56.6%
Mean: 73.8%
Upper Quartile: 96.6%
Maximum: 100%
Quartile Range: 39.9%
Std Dev: 29.5%

Female
# of providers: 2214
# of patients: 178353
Minimum: 0.00%
Lower Quartile: 51.2%
Mean: 71.1%
Upper Quartile: 95.6%
Maximum: 100%
Quartile Range: 44.4%
Std Dev: 30.7%

Age: <60
# of providers: 2203
# of patients: 76319
Minimum: 0.00%
Lower Quartile: 50.0%
Mean: 70.1%
Upper Quartile: 95.7%
Maximum: 100%
Quartile Range: 45.7%
Std Dev: 31.3%

Age: 60 -< 70
# of providers: 2213
# of patients: 95585
Minimum: 0.00%
Lower Quartile: 54.5%
Mean: 72.6%
Upper Quartile: 97.3%
Maximum: 100%
Quartile Range: 42.8%
Std Dev: 31.1%

Age: 70 -< 80
# of providers: 2216
# of patients: 120341
Minimum: 0.00%
Lower Quartile: 58.4%
Mean: 74.4%
Upper Quartile: 97.9%
Maximum: 100%
Quartile Range: 39.6%
Std Dev: 30.3%

Age: >= 80
# of providers: 2217
# of patients: 112161
Minimum: 0.00%
Lower Quartile: 52.6%

Mean: 72.2%
Upper Quartile: 97.1%
Maximum: 100%
Quartile Range: 44.5%
Std Dev: 31.0%

Insurance: None
# of providers: 53
# of patients: 236
Minimum: 0.00%
Lower Quartile: 0.00%
Mean: 67.2%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 100%
Std Dev: 45.5%

Insurance: Private
# of providers: 586
# of patients: 39659
Minimum: 0.00%
Lower Quartile: 54.2%
Mean: 73.7%
Upper Quartile: 99.1%
Maximum: 100%
Quartile Range: 45.0%
Std Dev: 30.5%

Insurance: Medicaid
# of providers: 342
# of patients: 9930
Minimum: 0.00%
Lower Quartile: 50.0%
Mean: 69.3%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 50.0%
Std Dev: 33.7%

Insurance: Medicare
# of providers: 24
# of patients: 112
Minimum: 0.00%
Lower Quartile: 20.0%
Mean: 63.1%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 80.0%
Std Dev: 40.7%

Insurance: Other
# of providers: 55
# of patients: 116
Minimum: 0.00%
Lower Quartile: 0.00%
Mean: 65.4%
Upper Quartile: 100%

Maximum: 100%
Quartile Range: 100%
Std Dev: 42.8%

Race: White
# of providers: 1248
# of patients: 186216
Minimum: 0.00%
Lower Quartile: 58.0%
Mean: 74.2%
Upper Quartile: 97.5%
Maximum: 100%
Quartile Range: 39.4%
Std Dev: 30.4%

Race: Black
# of providers: 1070
# of patients: 14849
Minimum: 0.00%
Lower Quartile: 50.0%
Mean: 72.1%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 50.0%
Std Dev: 35.9%

Race: Other
# of providers: 761
# of patients: 4932
Minimum: 0.00%
Lower Quartile: 50.0%
Mean: 71.7%
Upper Quartile: 100%
Maximum: 100%
Quartile Range: 50.0%
Std Dev: 38.4%

The 2013 National Healthcare Disparities Report showed that disparities in care for heart failure exist across populations. [1]

Based on data collected by AHRQ, from 2001 to 2010, the hospitalization rate for congestive heart failure decreased significantly overall and for each racial/ethnic and area income group. Although the quality of care improved overall, the following trends were observed.

•In all years, Blacks had higher rates of admission for congestive heart failure compared with Whites while APIs had lower rates than Whites.

•In all years, residents of the highest area income quartile had lower rates than residents of the two lower area income quartiles.

•The 2008 top 4 State achievable benchmark for congestive heart failure admissions was 195 per 100,000 population. At current rates of improvement, Whites could achieve the benchmark in 6 years. APIs, Hispanics, and Blacks could achieve the benchmark in 1, 4, and 9 years, respectively. Residents of the lowest income quartile would need 13 years while residents of other income quartiles could achieve the benchmark in 7 years.

[1] Agency for Healthcare Research and Quality. 2013 National Healthcare Disparities Report.

http://www.ahrq.gov/research/findings/nhqrdr/nhdr13/2013nhdr.pdf. Published May 2014.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

It is estimated 5.1 million Americans >=20 years of age have heart failure. [1] Projections show that the prevalence of HF will increase 46% from 2012 to 2030, resulting in >8 million people >=18 years of age with HF [2]. In 2012, total cost for HF was estimated to be $30.7 million. Of this total, 68% was attributable to direct medical costs. [2] Projections show that by 2030, the total cost of HF will increase almost 127% to $69.7 billion from 2012. This equals about $244 for every US adult.[2]

On the basis of data from the community surveillance component of the ARIC study of the NHLBI, it is estimated that there are 825,000 new HF cases annually and that at ages <75 years, HF incidence is higher in blacks than whites.[1]

The annual rates per 1000 population of new HF events for white men are 15.2 for those 65 to 74 years of age, 31.7 for those 75 to 84 years of age, and 65.2 for those >=85 years of age. For white women in the same age groups, the rates are 8.2, 19.8, and 45.6, respectively. For black men, the rates are 16.9, 25.5, and 50.6 (unreliable estimate), and for black women, the estimated rates are 14.2, 25.5, and 44.0 (unreliable estimate), respectively (CHS, NHLBI). [3]

The death rate for those with heart failure remains high with about 50% of people diagnosed with HF will die within 5 years.13,19 [4,5]

There is data from multiple studies also showing the impact of Heart Failure on American. Some of this data is briefly summarized below.

Data from Olmsted County, MN, indicate the following:
-Among asymptomatic individuals, the prevalence of left ventricular diastolic dysfunction was 21% for mild diastolic dysfunction and 7% for moderate or severe diastolic dysfunction. The prevalence of systolic dysfunction was 6%. The presence of any left ventricular dysfunction (systolic or diastolic) was associated with an increased risk of developing overt HF, and diastolic dysfunction was predictive of all-cause death.[6] After 4 years of follow-up, the prevalence of diastolic dysfunction increased to 39.2%. Diastolic dysfunction was associated with development of HF during 6 years of subsequent follow-up after adjustment for age, hypertension, DM, and CAD (HR, 1 .81; 95% CI, 1.01–3.48). [7]

-Among individuals with symptomatic HF, the prevalence of left ventricular diastolic dysfunction was 6% for mild diastolic dysfunction and 75% for moderate or severe diastolic dysfunction.[8]The proportion of people with HF and preserved EF increased over time. Survival improved over time among individuals with reduced EF but not among those with preserved EF.[9]

In MESA, African Americans had the highest risk of developing HF, followed by Hispanic, white, and Chinese Americans (4.6, 3.5, 2.4, and 1.0 per 1000 person-years, respectively). This higher risk reflected differences in the prevalence of hypertension, DM, and socioeconomic status. African Americans had the highest proportion of incident HF not preceded by clinical MI (75%).[10]

In the NHLBI's ARIC study, the age-adjusted incidence rate per 1000 person-years was 3.4 for white women, less than for all other groups, that is, white men (6.0), black women (8.1), and black men (9.1). The 30-day, 1-year, and 5-year case fatality rates after

hospitalization for HF were 10.4%, 22%, and 42.3%, respectively. Blacks had a greater 5-year case fatality rate than whites (P<0.05). HF incidence rates in black women were more similar to those of men than of white women. The greater HF incidence in blacks than in whites is explained largely by blacks' greater levels of atherosclerotic risk factors.[11]

Data from Kaiser Permanente indicated an increase in the incidence of HF among the elderly, with the effect being greater in men. [12]

1c.4. Citations for data demonstrating high priority provided in 1a.3

[1] Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, Dai S, Ford ES, Fox CS, Franco S, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Huffman MD, Judd SE, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Mackey RH, Magid DJ, Marcus GM, Marelli A, Matchar DB, McGuire DK, Mohler ER 3rd, Moy CS, Mussolino ME, Neumar RW, Nichol G, Pandey DK, Paynter NP, Reeves MJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Wong ND, Woo D, Turner MB; on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2014 update: a report from the American Heart Association. Circulation. 2014;129:e28–e292.

[2] Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, Ikonomidis JS, Khavjou O, Konstam MA, Maddox TM, Nichol G, Pham M, Piña IL, Trogdon JG; on behalf of the American Heart Association Advocacy Coordinating Committee; Council on Arteriosclerosis, Thrombosis and Vascular Biology; Council on Cardiovascular Radiology and Intervention; Council on Clinical Cardiology; Council on Epidemiology and Prevention; Stroke Council. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail. 2013;6:606–619.

[3] Incidence and Prevalence: 2006 Chart Book on Cardiovascular and Lung Diseases. Bethesda, MD: National Heart, Lung, and Blood Institute; 2006.

[4] Roger VL, Weston SA, Redfield MM, et al. Trends in heart failure incidence and survival in a community-based population. JAMA. 2004;292:344-350.

[5]Murphy SL, Xu JQ, Kochanek KD. Deaths: final data for 2010. National Vital Statistics Report. Vol 61, No 4. Hyattsville, MD: National Center for Health Statistics; 2013.

[6] Redfield MM, Jacobsen SJ, Burnett JC Jr, Mahoney DW, Bailey KR, Rodeheffer RJ. Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. JAMA. 2003;289:194–202.

[7] Kane GC, Karon BL, Mahoney DW, Redfield MM, Roger VL, Burnett JC Jr, Jacobsen SJ, Rodeheffer RJ. Progression of left ventricular diastolic dysfunction and risk of heart failure. JAMA. 2011;306:856–863.

[8] Bursi F, Weston SA, Redfield MM, Jacobsen SJ, Pakhomov S, Nkomo VT, Meverden RA, Roger VL. Systolic and diastolic heart failure in the community. JAMA. 2006;296:2209–2216.

[9] Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. N Engl J Med. 2006;355:251–259.

[10] Bahrami H, Kronmal R, Bluemke DA, Olson J, Shea S, Liu K, Burke GL, Lima JA. Differences in the incidence of congestive heart failure by ethnicity: the Multi-Ethnic Study of Atherosclerosis. Arch Intern Med. 2008;168:2138–2145.

[11] Loehr LR, Rosamond WD, Chang PP, Folsom AR, Chambless LE. Heart failure incidence and survival (from the Atherosclerosis Risk in Communities study). Am J Cardiol. 2008;101:1016–1022.

[12] Barker WH, Mullooly JP, Getchell W. Changing incidence and survival for heart failure in a well-defined older population, 1970-1974 and 1990- 1994. Circulation. 2006;113:799–805.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*

Not Applicable.

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

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.)*
http://www.ama-assn.org/ama1/pub/upload/mm/pcpi/hfset-12-5.pdf

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
For endorsement maintenance, we are not including e-measure specifications as we did in the previous submission cycle. We have updated the measure specifications to include the applicable CPT and ICD-10 codes.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients for whom the quantitative or qualitative results of a recent or prior (any time in the past) LVEF assessment is documented* within a 12 month period.

*Documentation must include documentation in a progress note of the results of an LVEF assessment, regardless of when the evaluation of ejection fraction was performed.

Qualitative results correspond to numeric equivalents as follows:
Hyperdynamic: corresponds to LVEF greater than 70%
Normal: corresponds to LVEF 50% to 70% (midpoint 60%)
Mild dysfunction: corresponds to LVEF 40% to 49% (midpoint 45%)
Moderate dysfunction: corresponds to LVEF 30% to 39% (midpoint 35%)
Severe dysfunction: corresponds to LVEF less than 30%

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Once during the measurement period.

S.6. **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, 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.
Left ventricular ejection fraction (LVEF) < 40% or documentation of severely or moderately depressed left ventricular systolic function (G8738)

OR

Left ventricular ejection fraction (LVEF) >=40% or documentation as normal or mildly depressed left ventricular systolic function (G8739)

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients aged 18 years and older with a diagnosis of heart failure.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
See 'Registry Supplemental Resources' attached in appendix field A.1 for data dictionary and form.

Codes that are applicable to denominator are:
Diagnosis for heart failure (ICD-9-CM): 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9

Diagnosis for heart failure (ICD-10-CM): I11.0, I13.0, I13.2, I50.1, I50.20, I50.21, I50.22, I50.23, I50.30, I50.31, I50.32, I50.33, I50.40, I50.41, I50.42, I50.43, I50.9

Patient encounter during the reporting period (CPT): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99304, 99305, 99306, 99307, 99308, 99309, 99310, 99324, 99325, 99326, 99327, 99328, 99334, 99335, 99336, 99337, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
None.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Not Applicable.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not Applicable.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*

Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. Type of score:
Rate/proportion
If other:

S.17. 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.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*
To calculate performance rates:

1) Find the patients who meet the initial patient 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 patient 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 patient population and denominator are identical.

4) From the patients within the denominator, find the patients who qualify for the Numerator (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.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not Applicable. The measure is not based on a sample.

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable. This measure is not based on a survey.

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If data required to determine if an individual patient should be included in a specific performance measure based on defined criteria is missing, those cases would ineligible for inclusion in the denominator and therefore the case would be deleted. If data required to determine if a denominator eligible patient qualifies for the numerator (or has a valid exclusion/exception) is missing, this case would represent a quality failure.

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Registry

S.24. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument e.g. name of database,*

clinical registry, collection instrument, etc.)
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is currently being used in the ACCF PINNACLE registry for the outpatient office setting. See attached form and data dictionary.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **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.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
HF_LVEF_Assessment_0079_Testing_Form_Version_6.5_6.22.15.pdf

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0079
Measure Title: Heart Failure: Left Ventricular Ejection Fraction Assessment (Outpatient Setting)
Date of Submission: 6/23/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form

refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The primary analysis was conducted at the level of the individual provider and included all patients with heart failure (HF) cared for by that provider and captured in the PINNACLE Registry during the one-year study period. The PINNACLE Registry systematically maps each practice's Electronic Health Record to the data elements required for the Registry, with careful validation of the translation process prior to enrollment and reporting the results back to the practice. Using these data, we were able to calculate the number of patients who should have undergone a left ventricular ejection fraction (LVEF) assessment during the 12-month period, or a strong reason why the LVEF assessment was not performed is documented. This means that every patient in that provider's practice is included. For this measure, providers with less than 10 eligible patient encounters during the study period were excluded, since estimates of reliability are unstable with such small numbers. All other cases from all practices and providers were included. We included all visits for each patient in these analyses and meeting the performance measure on any single visit within the year met the criterion for this measure.

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

The primary analysis included encounters between 01/01/2014-12/31/2014. Additionally, we used data from 01/01/2013 thru 12/31/2013 for temporal comparison.

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.26***)	**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*)

2013

2,254 providers met the minimum number of eligible patients (10) for inclusion in the reliability analysis. The average number of eligible patients for providers included is 181.6 for a total of 409,332 patients. The range of number of patients for providers included is from 10 to 2,730. As noted above, providers with fewer than 10 eligible patient encounters during the study period were excluded.

The unit of analysis for this measure is the provider. A description of the providers studied for the 2013 calendar year is shown below:

	Total **n = 2254**

	Total n = 2254
Provider gender	
(1) Male	1819 (80.8%)
(2) Female	433 (19.2%)
Missing (.)	2
Provider categories	
NP/PA	220 (9.9%)
MD/DO	1951 (88.0%)
RN/nurses	47 (2.1%)
Missing (.)	36
Region	
(1) Northeast	335 (14.9%)
(2) Midwest	691 (30.7%)
(3) South	843 (37.4%)
(4) West	385 (17.1%)

2014

2,219 providers met the minimum number of eligible patients (10) for inclusion in the reliability analysis. The average number of eligible patients for providers included is 182.2 for a total of 404,406 patients. The range of numbers of patients for providers included is from 10 to 2,990. As noted above, providers with fewer than 10 eligible patient encounters during the study period were excluded.

The unit of analysis for this measure is the provider. A description of the providers studied for the 2014 calendar year is shown below:

	Total n = 2219
Provider gender	
(1) Male	1772 (79.9%)
(2) Female	447 (20.1%)
Provider categories	
NP/PA	228 (10.5%)
MD/DO	1907 (87.5%)
RN/nurses	45 (2.1%)
Missing (.)	39
Region	
(1) Northeast	334 (15.1%)
(2) Midwest	543 (24.5%)
(3) South	931 (42.0%)
(4) West	411 (18.5%)

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

2013

There are a total of 409,332 patients included in the temporal comparison that were treated in 2013. Patients' characteristics are provided below:

	Total
	n = 409332
Race	
(1) White	202768 (89.5%)
(2) Black	19665 (8.7%)
(3) Other	4211 (1.9%)
Missing (.)	90895
Insurance	
(0) No insurance	311 (0.6%)
(1) Private	33938 (70.5%)
(2) Medicare	13596 (28.2%)
(3) Medicaid	107 (0.2%)
(4) Other	204 (0.4%)
Missing (.)	51502
Age	
18 to <60	77884 (19.0%)
60 to <70	96129 (23.5%)
70 to <80	120506 (29.4%)
80 to 114	114813 (28.0%)
Sex	
(1) Male	228280 (55.8%)
(2) Female	180550 (44.2%)
Missing (.)	300
BMI (kg/m2)	30.6 ± 8.4
Missing	86645
Diabetes Mellitus	122242 (32.2%)
Coronary Artery Disease	241431 (66.8%)
Hypertension	312835 (86.5%)
Atrial Fibrillation/Flutter	151755 (40.4%)
Peripheral Arterial Disease	56306 (15.4%)
Stroke/TIA	19269 (7.4%)
Myocardial Infarction	88104 (25.8%)

2014

There are a total of 404,406 patients included in the primary analysis (2014), whose characteristics are described below:

	Total
	n = 404406
Race	
(1) White	186216 (90.4%)
(2) Black	14849 (7.2%)
(3) Other	4932 (2.4%)
Missing (.)	61304
Insurance	
(0) No insurance	236 (0.5%)
(1) Private	39659 (79.2%)
(2) Medicare	9930 (19.8%)
(3) Medicaid	112 (0.2%)
(4) Other	116 (0.2%)
Missing (.)	57545
Age	
18 to <60	76319 (18.9%)
60 to <70	95585 (23.6%)
70 to <80	120341 (29.8%)
80 to 114	112161 (27.7%)
Sex	
(1) Male	224353 (55.7%)
(2) Female	178353 (44.3%)
Missing (.)	12482
BMI (kg/m2)	30.6 ± 8.2
Missing	74286
Diabetes Mellitus	118503 (30.8%)
Coronary Artery Disease	235361 (64.4%)
Hypertension	309664 (86.0%)
Atrial Fibrillation/Flutter	160293 (41.5%)
Peripheral Arterial Disease	54749 (15.4%)
Stroke/TIA	21525 (7.4%)
Myocardial Infarction	84597 (24.2%)

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

The dataset described above was used for all aspects of testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We do not currently collect any of the SDS variables examples listed above. As is noted in other sections of this testing form we do collect data or race as well as insurance type.

2a2. RELIABILITY TESTING

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

Reliability 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], where the latter represents the within-physician estimate of our error in assessing their 'true' performance. Thus, this assessment of 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 five different distributions of provider volumes: at the minimum number of quality reporting events for the measure; at the mean number of quality reporting events per physician; and at the 25th, 50th and 75th percentiles of the number of quality reporting events.

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*)

2013 – In 2013, the signal-noise ratios are shown below:

Description	Number of Patients	Signal-to-Noise Ratio
Minimum	10	0.988
25th percentile	61	0.994
50th percentile	128	0.996
75th percentile	227	0.998
Average	182	0.997

2014 – In 2014, the signal-noise ratios are shown below:

Description	Number of Patients	Signal-to-Noise Ratio
Minimum	10	0.989
25th percentile	66	0.994
50th percentile	130	0.996
75th percentile	217	0.997
Average	183	0.997

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?*)

For this measure the reliability was very high and was similar for 2013 and 2014, supporting the reproducibility of these estimates across years. At the minimum number of patient visits required (>10) the average reliability was 0.988 and 0.989 for 2013 and 2014, respectively. For providers with the median number of patient encounters, the reliability was even higher, 0.997 in both years. Given that a reliability of 0.70 is generally considered a minimum threshold for acceptability, and 0.80 is considered very good reliability, these data suggest that the measure is exceedingly good at describing true differences across physicians.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Clinical Evidence: Evaluation of LVEF in patients with heart failure provides important information that is required to appropriately direct treatment. Several pharmacologic therapies have demonstrated efficacy in slowing disease progression and improving outcomes in patients with left ventricular systolic dysfunction. LVEF assessed during the initial evaluation of patients presenting with heart failure can be considered valid unless the patient has demonstrated a major change in clinical status, experienced or recovered from a clinical event, or received therapy that might have a significant effect on cardiac function. Both journal articles that are referenced in Section 1a7.9 of the evidence form supports the assessment of ejection fraction to enable appropriate guideline-directed medical treatment.

Construct validity was difficult to establish because there has not been an independent audit of these data. However, it is important to note that an independent audit would merely involve an abstractor reviewing the same medical record from which PINNACLE directly abstracts its data and, given the identical source of the data, any error observed would either be due to the auditor incorrectly abstracting the data from the EHR or

PINNACLE incorrectly mapping the data elements from the EHR. To address the latter, we conduct detailed analyses to insure that this does not happen and quarantine (i.e. not report) data that fails our addition Data Quality Review process. Validity of measure data elements in PINNACLE is routinely evaluated on a quarterly basis as part of the standard data extraction and analytic data set creation process. First, all relevant data elements are reviewed at the record level to ensure that individual data values are valid; any invalid values are set to missing. Next, the *distribution* of each data element is reviewed, aggregating both across practices and across calendar quarters within each practice, to identify outliers, suspicious patterns and/or systematic changes in the prevalence of the data element that may suggest data mapping errors or unanticipated changes in definitions, coding consistency, data completeness, etc. Identification of suspicious data includes both statistical criteria, using quality control charts with rigorous definitions of "out of control" rates, and manual clinical review of each distribution for plausibility. Records that are flagged as suspicious by these criteria are quarantined and excluded from analysis and reporting. In 2013 the rate of records not passing the quality evaluation was 3.1% and in 2014 it was 8.7%. Feedback reports are generated to facilitate investigation of data issues at the practice level to verify accuracy of abstraction and to remap elements whose definitions or recording have changed.

Content validity for this measure was assessed by expert work group members during the development process. Additional input on the content validity of draft measures was established through a 30-day public comment period and concurrent formal peer review process. All comments received were reviewed by the expert work group and the measures were adjusted as needed. Additionally, the measure underwent review and approval by the Board of Trustees of the ACC and the Science Advisory and Coordinating Committee of the AHA, as well as review and voting by the PCPI membership. Members of the expert work group that developed the measure included: Robert O. Bonow, MD, MACC, FAHA, FACP (Co-Chair) (cardiology); Theodore G. Ganiats, MD (Co-Chair) (family medicine; measure methodology); Craig T. Beam, CRE (patient representative); Kathleen Blake, MD (cardiac electrophysiology); Donald E. Casey, Jr., MD, MPH, MBA, FACP (internal medicine); Sarah J. Goodlin, MD (geriatrics, palliative medicine); Kathleen L. Grady, PhD, APN, FAAN, FAHA (cardiac surgery); Randal F. Hundley, MD, FACC (cardiology, health plan representative); Mariell Jessup, MD, FACC, FAHA, FESC (cardiology, heart failure); Thomas E. Lynn, MD (family medicine, measure implementation); Frederick A. Masoudi, MD, MSPH (cardiology); David Nilasena MD, MSPH, MS (general preventive medicine, public health, measure implementation); Paul D. Rockswold, MD, MPH (family medicine); Ileana L. Piña, MD, FACC (cardiology, heart failure); Lawrence B. Sadwin (patient representative); Joanna D. Sikkema, MSN, ANP-BC, FAHA (cardiology); Carrie A. Sincak, PharmD, BCPS (pharmacy); John Spertus, MD, MPH (cardiology); Patrick J. Torcson, MD, FACP, MMM (hospital medicine); Elizabeth Torres, MD (internal medicine); Mark V. Williams, MD, FHM (hospital medicine); John B Wong, MD (internal medicine).

Face validity of the measure score was systematically assessed as follows:

After the measure was fully specified, members of two existing committees, one at the ACC and one at AHA, with expertise in in general cardiology, interventional cardiology, heart failure, electrophysiology and quality improvement, outcomes research, informatics and performance measurement, who were not involved in development of the measure, were asked to review the measure specifications and 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."

The respondents recorded their rating on a scale of 1-5, where 1= Strongly Disagree; 3=Neither Agree nor Disagree; 5= Strongly Agree

Forty two (42) committee members completed the survey and provided a mean importance rating of 4.24, with 85.7% agreeing with the use of the measure for quality assessment.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Additionally, the results of the expert panel rating of the validity statement were as follows:

N = 42; Mean rating = 4.24 and 85.7% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality

Frequency Distribution of Ratings
1 - <2> (Strongly Disagree)
2 - <3>
3 - <1> (Neither Agree nor Disagree)
4 - <13>
5 - <23> (Strongly Agree)

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The measure was judged to have high face validity by both its clinical importance and by the group of experts asked to rate it. The majority of experts agreed that the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

2b3. EXCLUSIONS ANALYSIS
NA ☒ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Since not all patients with heart failure, in rare circumstances, might not meet the guideline recommendations for EF evaluation, exclusions in this measure are intended to remove patients for whom an EF assessment might not be appropriate (e.g. patient refusal). We divide these into two categories: Exclusions and Exceptions. Exclusions arise when patients who are included in the initial patient or eligible population for the measure set do not meet the denominator criteria specific to the intervention required by the numerator. There are no known exclusions for this assessment, which are usually derived from evidence-based guidelines. Exceptions are not absolute, and are based on clinical judgment and individual patient characteristics, such as refusal to participate in the test.

2b3.2. 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*)

There are no exclusions and there were no exceptions reported for this measure in any patients.

2b3.3. 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)

All patients are eligible for this measure and all providers caring for HF patients should be able to have a performance rate calculated on their entire population.

_____

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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)

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

We examined variation in provider performance on this measure based on sex, age, race and a number of other patient factors to identify variations. The findings fare represented for 2013 and 2014 respectively.

2013

Label	# of providers	# of patients	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
Male	2250	228280	0.00%	47.6%	69.3%	95.5%	100%	47.8%	31.8%
Female	2250	180550	0.00%	41.0%	66.2%	94.4%	100%	53.4%	32.8%
Age: <60	2236	77884	0.00%	42.0%	66.1%	94.4%	100%	52.4%	33.3%
Age: 60 -< 70	2246	96129	0.00%	46.3%	68.4%	96.4%	100%	50.1%	33.0%
Age: 70 -< 80	2253	120506	0.00%	46.8%	69.4%	96.9%	100%	50.1%	32.8%
Age: >= 80	2243	114813	0.00%	42.9%	67.3%	96.0%	100%	53.1%	33.4%
Insurance: None	69	311	0.00%	25.0%	69.1%	100%	100%	75.0%	42.7%
Insurance: Private	596	33938	0.00%	33.3%	64.4%	97.1%	100%	63.8%	35.7%
Insurance: Medicaid	382	13596	0.00%	28.6%	60.0%	90.2%	100%	61.6%	35.6%
Insurance: Medicare	23	107	0.00%	0.00%	47.6%	91.7%	100%	91.7%	44.2%
Insurance: Other	76	204	0.00%	50.0%	76.3%	100%	100%	50.0%	36.3%
Race: White	1479	202768	0.00%	50.8%	71.2%	96.9%	100%	46.1%	31.4%
Race: Black	1267	19665	0.00%	48.6%	70.5%	100%	100%	51.4%	36.3%
Race: Other	842	4211	0.00%	33.3%	67.2%	100%	100%	66.7%	40.6%

2014

label	# of providers	# of patients	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
Male	2214	224353	0.00%	56.6%	73.8%	96.6%	100%	39.9%	29.5%
Female	2214	178353	0.00%	51.2%	71.1%	95.6%	100%	44.4%	30.7%

label	# of providers	# of patients	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
Age: <60	2203	76319	0.00%	50.0%	70.1%	95.7%	100%	45.7%	31.3%
Age: 60 -< 70	2213	95585	0.00%	54.5%	72.6%	97.3%	100%	42.8%	31.1%
Age: 70 -< 80	2216	120341	0.00%	58.4%	74.4%	97.9%	100%	39.6%	30.3%
Age: >= 80	2217	112161	0.00%	52.6%	72.2%	97.1%	100%	44.5%	31.0%
Insurance: None	53	236	0.00%	0.00%	67.2%	100%	100%	100%	45.5%
Insurance: Private	586	39659	0.00%	54.2%	73.7%	99.1%	100%	45.0%	30.5%
Insurance: Medicaid	342	9930	0.00%	50.0%	69.3%	100%	100%	50.0%	33.7%
Insurance: Medicare	24	112	0.00%	20.0%	63.1%	100%	100%	80.0%	40.7%
Insurance: Other	55	116	0.00%	0.00%	65.4%	100%	100%	100%	42.8%
Race: White	1248	186216	0.00%	58.0%	74.2%	97.5%	100%	39.4%	30.4%
Race: Black	1070	14849	0.00%	50.0%	72.1%	100%	100%	50.0%	35.9%
Race: Other	761	4932	0.00%	50.0%	71.7%	100%	100%	50.0%	38.4%

2b5.2. 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*)

Provided below is the testing that identified the differences in performance measure scores for 2013 and 2014.

2013

Overall mean performance on this measure is 67.8%, with a standard deviation of 32.0%. The minimum score equals 0.00%, while the maximum score equals 100.00%. The interquartile score is equal to 49.6%.

2,254 providers were measured, and the patient study sample equals 409,332. 55.8% of the sample is male. 89.5% of the sample is white, 8.7% is black, and 1.9% identified as "other." The sample reached across all US regions, with 14.9% of providers in the Northeast, 30.7% of providers in the Midwest, 37.4% of providers in the South, and 17.1% of providers in the West.

# of providers	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
2254	0.00%	45.2%	67.8%	94.8%	100%	49.6%	32.0%

	Mean
Decile 1	4.0%
Decile 2	23.5%
Decile 3	44.5%

	Mean
Decile 4	62.3%
Decile 5	75.5%
Decile 6	84.9%
Decile 7	90.6%
Decile 8	94.7%
Decile 9	97.7%
Decile 10	99.7%

2014

Overall mean performance on this measure is 72.5%, with a standard deviation of 29.9%. The minimum score equals 0.00%, while the maximum score equals 100.00%. The interquartile score is equal to 41.6%.

2,219 providers were measured, and the patient study sample equals 404,406. 55.7% of the sample is male. 90.4% of the sample is white, 7.2% is black, and 2.4% identified as "other." The sample reached across all US regions, with 15.1% of providers in the Northeast, 24.5% of providers in the Midwest, 42.0% of providers in the South, and 18.5 % of providers in the West.

# of providers	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
2219	0.00%	54.3%	72.5%	95.9%	100%	41.6%	29.9%

	Mean
Decile 1	7.2%
Decile 2	34.7%
Decile 3	54.4%
Decile 4	69.9%
Decile 5	82.1%

	Mean
Decile 6	89.4 %
Decile 7	93.1 %
Decile 8	95.9 %
Decile 9	98.0 %
Decile 10	99.7 %

2b5.3. 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?*)

2013: A moderate to large amount of variability was noted among providers. The performance-met rate range was 0-100% with the inter-quartile range being 45.2% to 94.8%. This yielded a Median Rate Ratio of 2.13 (2.07, 2.18). The Median Rate Ratio measures the variation across providers for statistically 'identical' patients and suggests that a patient presenting to 1 provider, as opposed to another, would, on average, be 2.3-times more likely to have their EF assessed.

2014: A moderate, but slightly lower, amount of variability was noted among providers. The performance-met rate range was 0-100% with the inter-quartile range being 54.3% to 95.9%. This yielded a Median Rate Ratio of 1.92 (1.88, 1.97).

2b6. 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 SDS 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 SDS 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.***

2b6.1. 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*)

2b6.2. 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*)

2b6.3. 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

In PINNACLE, if a data field is not fulfilled, it is assumed that the process of care was not done. Thus, there are no missing values for this measure, although it is conceivable that the physician did assess the LVEF of the patient, but did not record them in the electronic health record, form which the data in PINNACLE is directly abstracted. If there are missing fields, we believe that these will be rapidly corrected on there is physician level accountability for recording these data occurs.

2b7.2. 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*)

Given our assumptions, noted above, we did not conduct an empirical analysis of frequency or distribution of missing data. For this measure, missing data is reported as a quality failure.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

We do not believe any biases are introduced in the assessing of individual physician performance and continued endorsement of this measure would lead to improved care.

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)

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)*

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM 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)*.

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, the ACC, the AHA, the PCPI and its members disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT®) or other coding contained in the specifications.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement (Internal to the specific organization)	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) ACC Pinnacle Registry URL: http://www.ncdr.com/webncdr/pinnacle/

4a.1. **For each CURRENT use, checked above, provide:**

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

PINNACLE Registry (URL: http://www.ncdr.com/webncdr/pinnacle/)
In 2008, the American College of Cardiology Foundation launched the PINNACLE program (Formerly known as the Improving Continuous Cardiac Care or IC3). The PINNACLE Registry® continues to grow rapidly, with more than 3446 providers representing almost 960 unique office locations across the U.S submitting data to the registry. As of March 2015, the registry has more than 19.8 million patient encounter records representing approximately 4.85 million patients. PINNACLE assists practices in understanding and improving care through the production and distribution of quarterly performance reports. These reports, covering all valid patient encounters, detail adherence to 28 cardiovascular clinical measures at the physician, location, and practice levels across coronary artery disease, hypertension, heart failure and atrial fibrillation. All jointly developed ACC/AHA/PCPI performance measures for these topics are reported by the registry.

4a.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 are continuously seeking opportunities to advocate for 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.

4a.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.*)
We are continuously seeking opportunities to advocate for expanded use of this measure in government or other programs, including those intended for accountability or public reporting.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

The mean performance rates from the Pinnacle registry increased from 2013 to 2014 from 67.8% to 72.5%. In 2013, 2254 providers were measured, and the patient study sample equals 409,332. In 2014, 2219 providers were measured, and the patient study sample equaled 404,406. The statistical significance of these results was not analyzed.

4b.2. **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 ACCF/AHA and PCPI create 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.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

We are not aware of any unintended consequences at this time, but we continuously monitor for them.

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)**

0135 : Evaluation of Left ventricular systolic function (LVS)

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

5a. Harmonization

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 completely harmonized?

No

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

This measure is inpatient based and focuses on the assessment occurring prior to discharge. Our measure looks at whether the assessment was performed during a 12 month period for a patient with a diagnosis of heart failure.

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.)

Related Measures: NQF # 0135: Evaluation of Left ventricular systolic function (LVS). This measure is inpatient based and focuses on

the assessment occurring prior to discharge. Our measure looks at whether the assessment was performed during a 12 month period for a patient with a diagnosis of heart failure.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** HF_LVEF_Assessment_0079_PINNACLE_Registry_data_collection_form_and_dictionary.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American College of Cardiology
Co.2 **Point of Contact:** Penelope, Solis, comment@acc.org, 202-375-6576-
Co.3 **Measure Developer if different from Measure Steward:** American College of Cardiology
Co.4 **Point of Contact:** Penelope, Solis, comment@acc.org, 202-375-6576-

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.
Robert O. Bonow, MD, MACC, FAHA, FACP (Co-Chair) (cardiology)
Theodore G. Ganiats, MD (Co-Chair) (family medicine; measure methodology)
Craig T. Beam, CRE (patient representative)
Kathleen Blake, MD (cardiac electrophysiology)
Donald E. Casey, Jr., MD, MPH, MBA, FACP (internal medicine)
Sarah J. Goodlin, MD (geriatrics, palliative medicine)
Kathleen L. Grady, PhD, APN, FAAN, FAHA (cardiac surgery)
Randal F. Hundley, MD, FACC (cardiology, health plan representative)
Mariell Jessup, MD, FACC, FAHA, FESC (cardiology, heart failure)
Thomas E. Lynn, MD (family medicine, measure implementation)
Frederick A. Masoudi, MD, MSPH (cardiology)
David Nilasena MD, MSPH, MS (general preventive medicine, public health, measure implementation)
Paul D. Rockswold, MD, MPH (family medicine)
Ileana L. Piña, MD, FACC (cardiology, heart failure)
Lawrence B. Sadwin (patient representative)
Joanna D. Sikkema, MSN, ANP-BC, FAHA (cardiology)
Carrie A. Sincak, PharmD, BCPS (pharmacy)
John Spertus, MD, MPH (cardiology)
Patrick J. Torcson, MD, FACP, MMM (hospital medicine)
Elizabeth Torres, MD (internal medicine)
Mark V. Williams, MD, FHM (hospital medicine)
John B Wong, MD (internal medicine)

These measures were developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study must be equal contributors to the measure development process. In addition, the we strived to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group.

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:** 12, 2010
Ad.4 **What is your frequency for review/update of this measure?** Every 3 years or as new evidence becomes available that materially affects the measures
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.6 **Copyright statement:** This Physician Performance Measurement Set (PPMS) and related data specifications were developed by the Physician Consortium for Performance Improvement® (the Consortium) including the American College of Cardiology (ACC), the American Heart Association (AHA) and the American Medical Association (AMA) to facilitate quality improvement activities by physicians. The performance measures contained in this PPMS are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. While copyrighted, they 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 performance measures for commercial gain, or incorporation of the performance measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the PPMS require a license agreement between the user and the AMA, (on behalf of the Consortium) or the ACC or the AHA. Neither the AMA, ACC, AHA, the Consortium nor its members shall be responsible for any use of this PPMS.

THE MEASURES 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, the ACC, the AHA, the Consortium and its members disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT®) or other coding contained in the specifications.

Ad.7 **Disclaimers:** See copyright statement above.

Ad.8 **Additional Information/Comments:** The ACCF, AHA, and PCPI have a formal measurement review process that stipulates regular (usually on a three-year cycle, when feasible) review of the measures. The process can also be activated if there is a major change in scientific evidence, results from testing or other implementation issues are noted that materially affect the integrity of the measure.

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 #: 0083
De.2. **Measure Title:** Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)
Co.1.1. **Measure Steward:** AMA-PCPI
De.3. **Brief Description of Measure:** Percentage of patients aged 18 years and older with a diagnosis of heart failure with a current or prior LVEF < 40% who were prescribed beta-blocker therapy either within a 12 month period when seen in the outpatient setting or at 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 hospital discharge

*Prescribed may include:
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
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

**Beta-blocker therapy should include bisoprolol, carvedilol, or sustained release metoprolol succinate. (see technical specifications for additional information on medications)
S.7. **Denominator Statement:** All patients aged 18 years and older with a diagnosis of heart failure with a current or prior LVEF < 40%

LVEF < 40% corresponds to qualitative documentation of moderate dysfunction or severe dysfunction
S.10. **Denominator Exclusions:** 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)
Documentation of patient reason(s) for not prescribing beta-blocker therapy
Documentation of system reason(s) for not prescribing beta-blocker therapy

De.1. **Measure Type:**
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry
S.26. **Level of Analysis:** Clinician : Group/Practice, Clinician : Individual
Is this an eMeasure? ☒ **Yes** ☐ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☒ **Yes** ☐ **No**

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: 8/10/09 **Most Recent Endorsement Date:** 1/18/12

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

Public and Member Comments

- Concerns about broad exclusions.
- Clarification requested regarding the setting and data collection for this measure.

Developer response:

- This is a clinician-level measure for the outpatient setting.
- These measures have been tested and found to be generally feasible in EHR, paper, and claims data sources.

Steering Committee: Reviewed comments and developer's responses. No change to recommendations.

IF this measure is paired/grouped, NQF#/title: 0081: eart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)

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) Therapy for Left Ventricular Systolic Dysfunction) address related aspects of care for effective treatment for patients with heart failure and should be measured concurrently. Both ACE inhibitors and beta-blockers have been shown to reduce mortality and hospitalizations and improve a patient's clinical status. ARBs can be considered a reasonable alternative for ACE inhibitors. Combined treatment with these agents 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.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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.

The developer provides the following evidence for this process measure:

- This clinician-level process *registry* and *eMeasure* calculates the percentage of patients aged 18 years and older with a diagnosis of heart failure with a current or prior LVEF < 40% who were prescribed beta-blocker therapy either within a 12 month period when seen in the outpatient setting or at hospital discharge.
- The developer provides the 2013 ACCF/AHA guideline for the management of heart failure (Class I, Level A) that recommends the use of 1 of the 3 beta blockers proven to reduce mortality (eg, bisoprolol, carvedilol, and sustained release metoprolol succinate) for all patients with current or prior symptoms of heart failure, unless contraindicated.
- The articles supporting the beta-blocker recommendation were from 1989-2009 but the developers report that the overall literature search was through October 2011 with select articles included through April 2013.
- QQC: 17 Randomized Controlled Trials, 3 comparative studies.
- The developers provide 4 studies from 2014-2015. The additional studies recommended including nebivolol; compared cardevilol to metoprolol and bisoprolol; and looked at the use of beta-blockers in patients with heart failure and atrial fibrillation – the developers state that they will wait for revised guideline recommendations before considering changes to the measure.

- The developer provides a diagram of how the initiation of beta-blocker therapy for patients with heart failure or LVEF < 40% is linked to patient outcomes.

Questions for the Committee:

- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- The developers provided the average performance rates from the PQRS Experience Report from 2010-2013 and data from the literature that shows a beta-blocker was prescribed for 11,868 (86.2%) of 13,772 eligible patients Additional data (std dev, min, max, interquartile range, scores by decile, number of measured entities) that is required per NQF policy for maintenance is not provided.
- The developers report that disparities data from the federal reporting programs using this measure have not yet been made available for them to analyze and report.
- The developers provide data from a 2011 study that suggests that there are some racial and ethnic disparities in the receipt of pharmacological therapy for CHF among TRICARE beneficiaries.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - The evidence to support the measure focus was established through several clinical trials. This measure was established as a non-EHR measure in the past. The evidence source is the 2013 ACCF/ACA practice guidelines.
   - The developer provided ample evidence supporting the measure focus for this clinician-level process registry and eMeasure around beta blocker therapy for left ventricular systolic dysfunction (LVSD). The developer cited using the most current ACC/AHA guidelines that recommend as a Class IA the use of 1 of the 3 beta blockers proven to reduce mortality (eg bisoprolol, carvedilol, and sustained release metoprolol succinate) for all patients with current or prior symptoms of heart failure. The source also served as the systematic review of the body of the evidence.
   - As it relates to the quality, quantity and consistency of the evidence supporting this measure, it is quite high with 17 RCTs, and 3 comparative studies, representing literature from 1989 through 2009. However the developer also cited 4 additional recent studies conducted since the systematic review that included nebivolol; a comparison of metoptolol and bisoprolol; evaluated beta blockers in patients with AF and HF that was not included to support any measure changes. The developer will await updated guideline recommendations to refine the measure in any way based upon this new evidence.
   - Based upon the evidence provided, the evidence is directly applicable to the process of care being measured.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- This is a process measure that is a paired measure with 0081.
- The evidence for this measure shows that the measure is directly related to an improvement of clinical status, mortality and risk of hospitalization. The evidence is long-standing. The clinical practice guideline is the 2013 ACCF/AHA 2013 Practice Guidelines. Evidence is Level A – multiple RCTs in multiple populations, evidence from 1989-2009, with updates through 2013. Evidence included 17 randomized controlled trials and 3 comparative studies. Benefits were seen in those with and without diabetes, men and women, and in blacks.
- Harms were noted in the evidence review and those potential harms were included as medical exclusions from the measure. (bradycardia, heart block, hypotension, asthma).
- QQC provided

1b. ***Committee's Comments on Performance Gap:***

- Performance gap evidence is provided from 2010-2013 PQRS which shows a performance around 75-85% adherence without sustained improvement over the 4 years presented. This was consistent with results from the IMPROVE-HF registry.
- Disparities data have not been reported and aren't available at this time. A single study from 2011 using TRICARE data found the AA were less likely to receive treatment.
- Heart Failure affects nearly 6 million Americans... it continues to have huge epidemiologic and economic implications, thus remaining a high priority area within healthcare. Based upon the developers submission, there is still a gap in care that warrants a national performance measures. Average performance rates from the PQRS Experience Report from 2010-2013, combined with data from the literature suggests that BBlocker was prescribed 86.2% of eligible patients.
- As it relates to disparities, the developer noted that disparities data from the CMS has not been made available for analysis. That said, the developer provided reference to a study that suggest that there are some racial and ethnic disparities in the receipt of pharmacologic therapy for HF patients among TRICARE beneficiaries."

1c. ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure's data source is an EHR and/or registry. ICD 9, ICD 10 and CPT codes provided for the numerator and denominator for registry and inpatient setting. ICD-10 conversion methodology is not discussed. In both sets of specifications (registry and eMeasure), the logic is unambiguous. Higher scores equal better quality.
- In the outpatient setting, "prescribed" may include a beta-blocker prescription given to the patient at one or more visits in the measurement period or beta-blocker listed as current medication. In the inpatient setting, "prescribed" may include prescription given at discharge or beta-blocker listed in discharge medication list.
- The measure is intended for use in an office visit, outpatient consultation, nursing facility, long-term care residential facility, home health and provider interaction during the measurement period.
- Both specifications state, "Beta-blocker therapy should include bisoprolol, carvedilol, and sustained release metoprolol succinate", though an extensive list of beta blockaders are provided for the eMeasure.
- The denominator details state "In the outpatient setting, 2 or more encounters are required to establish the eligible professional has an existing relationship with the patient". Two visits could be added to the measure description or denominator to provide accuracy in measure calculation.
- For the registry specification, the denominator patient, medical & system exclusions/exceptions are reported using a single HCPCS code G8541, though in the paired measure 0081 they are reported using separate CPT-II codes for each of the 3 type of exclusions/exceptions.

- Denominator exclusions/exceptions include broad definitions for medical reason, patient reason and system reason for not prescribing beta-blocker therapy; examples of exceptions provided using CPT-II codes. The developer states that exceptions should only be considered when the numerator activity was not performed, that they are not uniformly relevant across measures, and that there must be a clear rationale to permit an exception for a medical, patient, or system reason. In the provided value sets, broadly defined and inappropriate patient, medical and system reason denominator exclusions include but are not limited to: medical reason: 216952002 failure in dosage (event), patient reason: 224187001 variable income (finding) & 266966009 family illness (situation), system reason:266756008 medical care unavailable (situation).
- Missing numerator data represents a quality failure.
- The calculation algorithm is included.
- The measure is not risk adjusted and SDS variables were not captured for the measure. The developer encourages users to provide collect data and stratify results by race, ethnicity, administrative sex, and payer consistent with CMS' Measures Management System Blueprint and recent national recommendations put forth by the IOM and NQF.
- For the eMeasure, the developer should clarify the following:
    - Atrioventricular Block and Cardiac Pacer in Situ are listed as exceptions in the eMeasure specification, though they are not included in the registry specifications. Should these be exclusions?
    - "Provider interactions" are listed as encounters in the value set spreadsheet and in the eMeasure specifications that include both face-to-face visits and non-face-to-face communications. The developer is encouraged to provide reasoning for inclusion, and clarify if all provider interactions are included in the denominator definition for a patient encounter.
    - "Communication: From Patient to Provider: Patient Reason for ACE Inhibitor or ARB Decline" is listed as an exception and is not included in the Registry specification. Please describe intent of this exception.
    - In the Initial Patient Population (IPP), "Encounter, Performed: Patient Provider Interaction" is listed, though it is not in the Data Criteria (QDM Variables).
    - "Encounter, Performed: Face-to-Face Interaction" is listed in the Data Criteria (QDM Variables), though not in the IPP.
- All eMeasure specifications and values sets meet current NQF eMeasure technical requirements and are provided on Sharepoint for SC review

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

The developers present 3 types of reliability testing:

- GPRO EHR Web-Interface – Performance measure score was performed using the CMS PQRS EHR Web Interface data base from January 2013-December 2013: 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.
    - 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. The developers conclude "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."

- GPRO Registry – Performance measure score was performed using the CMS PQRS GPRO database from January 2013-December 2013. The total number of physicians reporting on this measure is 1,748. Of those, 684 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, 39.1 percent of physicians are included in the analysis, and the average number of quality reporting events is 33.9 for a total of 23,175 events.
    - For this measure, the reliability at the minimum level of quality reporting events (10) was 0.86. The average number of quality reporting events for physicians included is 33.9. The reliability at the average number of quality reporting events was 0.96. The developers conclude "This measure has high reliability when evaluated at the minimum level of quality reporting events and high reliability at the average number of quality events."
- eMeasure – Critical data element (Validity against the Gold Standard) was conducted. Per NQF criteria, if empirical validity testing was performed of patient-level data, the rating from validity testing of patient-level data elements should be used.
    - The developer provides simple agreement results for critical data element validity. Comparison of the values for several data elements (electronic extracted vs. data abstracted) using EHR data was conducted and could satisfy the data element reliability criterion (Algorithm box 3; validity testing results described in 2b.2 below). Results were provided for most, but not all, critical data elements. Percent agreement statistics were presented; however, percent agreement does not adjust for agreement due to chance, and should not be used alone to demonstrate reliability.
    - Ideally, implementation of an eMeasure can be considered an automated process, and therefore the calculations will be consistent. The submitted eMeasure specification follows industry standards to represent the measure electronically which should enable automated data extraction and measure score calculation.
    - In addition to critical data element testing, the developer submitted pre-testing from the Measure Authoring Tool within the Bonnie Output that tests eMeasure logic. The measure logic successfully validated through the Bonnie Output.
- For both specifications, a missing data assessment was not performed. The developer states data missing from denominator excludes the patient from the measure, while missing numerator data counts as a measure "fail".
- Patient-level socio-demographic (SDS) variables were not captured as part of the testing.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The evidence states 3 beta-blockers should be used for this population, though a lengthy list of beta-blockades are provided for the eMeasure.
- The clinical practice guideline supporting this measure recommends the use of beta-blockers in patients with heart failure, which the specifications reflect.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

GPRO Registry Testing

- Validity of the measure score was assessed by systematic assessment of face validity by an expert panel of 12

members who either agreed or strongly agreed that this measure can accurately distinguish good and poor quality.

eMeasure Testing

- Data element validity testing against the gold standard was conducted to calculate parallel forms reliability for the measure. Test site was an academic general internal medical clinic; 254 charts were analyzed from 2007. The automated quality assessment results found:
  - a sensitivity of 100.0% for identifying patients with heart failure taking a beta-blocker,
  - 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 element validity testing was conducting by comparing, for several data elements, the values obtained from electronic extraction from 1 EHR to those obtained from those abstracted from the EHR by an abstractor. Simple agreement was provided for most, but not all, critical data elements. However, simple agreement does not adjust for agreement due to chance and thus should not be used alone to demonstrate validity; sensitivity/specificity statistics are preferred for demonstrating data element validity. Percentage agreement values were relatively high for most data elements considered. It appears that only one abstractor was utilized, which is acceptable for testing validity against the gold standard in an EHR.
- In addition to critical data element testing, the developer submitted pre-testing from the Measure Authoring Tool within the Bonnie Output that also tests eMeasure performance calculation. This testing does use "live" EHR patients, though NQF currently accepts Bonnie Output pre-testing when EHR testing was not provided. Results in the 56 "pre-test" patients demonstrated 100% agreement for identifying both expected and actual initial patient population, denominator, denominator exclusions, numerator, and denominator exceptions. Testing characteristics are provided for the 56 "pre-test" patients, with 95% of the of data elements concepts included in the Initial Patient Population (IPP), Denominator, Numerator and Denominator Exceptions, with 95% of all possible data elements covered in the pre-test sample.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- GPRO Registry: Of 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%. The types of exceptions reported are not available from the GRPO Registry.
  - EHR: 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. Atrioventricular Block and Cardiac Pacer in Situ are listed as exceptions in the eMeasure specification, though they are not included in the registry specifications.
- Broad exceptions (medical, patient & system) reasons are included within the provided data sets, and include non-relevant codes. For the Registry specification, one code encompasses all 3 types of exceptions.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This process measure is not risk adjusted.

2b5. Meaningful difference:

- 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).
- GPRO Registry: Based on the sample of 684 included physicians, the mean performance rate is 0.70, the median performance rate is 0.93 and the mode is 1.00. The standard deviation is 0.37. The range of the performance rate is 1.00, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.58 (0.42 - 1.00).

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- NQF criteria require a comparability assessment of data sources/methods, such as with multiple specifications for the same measure. This was not provided by the developer, as the eMeasure testing data used fictitious "pre-test" patients.
- For validity testing of the eMeasure, developers did find higher performance & detection rates for the records abstracted with EHR & manual reviews, than EHR-only reviews, and GPRO EHR Web-Interface also demonstrated higher performance than the GPRO Registry.

2b7. Missing Data

- Missing numerator data represent a quality failure.
- The developer reports "Data are not available to complete this testing."

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- Specifications – The denominator is patients 18 and older with a diagnosis of HF and an ejection fraction < 40%. The numerator is for patients prescribed beta blocker therapy within the past 12 months in either the outpatient or inpatient setting. BB therapy is limited to bisoprolol, carvedilol, or sustained release metoprolol succinate. It does not include nebivolol. It requires that the patient has at least 2 encounters with the provider in the measurement period to establish a patient-provider relationship.
- Exclusions – include standard AMA-PCPI exclusions for medical reasons, patient reasons, and system reasons. The patient reasons include a broad range of reasons including patient income, family situations and others. These reasons were problematic with CSAC in Phase 2 as CSAC these reasons should not be included, while the developer commented that these are standard across all of their measure sets, and the developer was not willing to adjust these standard exclusions for individual measures.
- As for the electronic specifications, all elements are specified with VSAC specifications. And are specified in the accepted standard of HQMF format, using the Quality Data Model (QDM) as required. It seemed as if heart block was included in the eMeasure specification as an exclusion but excluded from the registry specifications as an exclusion.
- It appears that if there is any missing data the case is excluded from measurement. The developer states this is a quality failure, but it is unclear exactly what this means.
- "This measure can be calculated using several data sources including registry and EHR. Measure specifications were provided for eMeasures, including the value sets. The developer noted that on an annual basis they review the supporting guidance to keep the measure adherent to current eCQM industry standards and still preserve the original intent of the measure.
- The numerator statement has been clearly defined, noting differences that may arise between the inpatient and outpatient settings. Moreover, the numerator specifications are explicit to just the three pharmacologic agents referenced in the guidelines and supported by the evidence. The time period for data collection was clear, again making note of any difference anticipated depending upon setting of care. In both cases, for EHR and for Registry, the definitions used were provided and/or included in the submission.
- I think that it is worthwhile to note that the measure truly seeks to compare apples to apples by requiring in the outpatient setting that the eligible provider must have had two or more encounters with the patient to

establish an existing relationship with the patient.

- The developer included the calculation algorithm in this submission. The measure is not risk adjusted nor were SDS variables captured.
- As noted, all eMeasure specifications and value sets meet current NQF requirements.
- As the NQF staff noted, there are some inconsistencies between the registry specifications and the eMeasure specifications. Perhaps some clarify regarding those differences would be helpful during the upcoming discussion."

2a2.: ***Committee's Comments on Reliability-Testing:***

- 3 types of reliability testing were provided: for web-interface, for registry use, and for EHRs, against a gold standard – presumably chart abstracted data from an academic setting. Emeasure testing was done using 5 different EHR systems. They looked at the exception analysis, signal to noise ratio.
- Signal to noise ratio showed a reliability of 87-96 depending on the data source. Overall they report moderate to high reliability depending on the data source.
- There are three different approaches used for reliability testing- GPRO EHR Web-interface, GPRO Registry, and eMeasure.
- The developer concluded based upon the GPRO EHR web-interface reliability testing from Jan 2013 through Dec 2013 that ""the measure has average reliability when evaluated at the minimum level of quality reporting events and high reliability at the average number of quality events."
- As for the GPRO Registry reliability testing, also conducted during the same time frame, the measure developer reported that ""the measure had high reliability when evaluated at the minimum level of quality reporting events and high reliability at the average number of quality events."
- Finally for the eMeasure reliability testing

2b1.: ***Committee's Comments on Validity-Specifications***:

- Consistent with the evidence.
- Specifications as outlined are not inconsistent with the evidence. There may be an opportunity to cross check the list of BBlockers included in the eMeasures specifications to ensure that included BBlockers are the Brand name associated with the three BBlockers included in the measure.

2b2.: ***Committee's Comments on Validity-Testing:***

- Face validity was done using an expert panel. HER testing against the gold standard detected 86% of patients on a BB and of the remaining patients, 1 in 3 had an exclusion criteria for an overall performance of 91%. Expert review identified 6additional patients for exclusion and only 1 exclusion that was identified to be false. This was an overall 12 of 18 patients with valid exclusion criteria (sensitivity 67%m and 1 of 13 patients with net exclusion criteria that did not have a valid exclusion out of a total of 254 patients.
- Face validity testing showed that with an N of 12 responses, there were 8 responses of Agree and 4 responses of strongly agreeing that the measure could distinguish between good and poor quality.
- Measure testing with the eMeasure found that measure exceptions were validated 95% of the time. Of 118 exceptions analyzed, 98% were medical reasons for not prescribing. Overall exception rate was 5%.
- Data elements were testing using the Bonnie Output, using 56 test patients. This demonstrated 100% performance with 95% coverage of data elements. Mean performance from the EHR data was 0.9 with a STD Dev 0.09 (IQR 0.85-0.97) and in the registry the mean was 0.7 with an STD Dev 0.37 (IQR 0.42-1)
- The measure is not risk adjusted.
- For GPRO Testing, face validity was assessed by an expert panel of 12 members. For Emeaure Testing, critical data element testing was conducted for several data elements utilizing one abstractor against the gold standard of the EHR. In addition, the measure developer submitted pre-testing form the Measure Authoring Tool to test eMeasure performance calculation.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Missing data could be a threat to validity although the developers do not discuss this. Missing data assessment was not done that I could tell.
- The measure is not risk adjusted.
- IN my review of the documentation, any threats to validity would arise due to the level of exception reporting. With EHR, measure exceptions were validated 95.3% of the time, 98% of those exceptions due to medical reasons that were consistent with clinical contraindications of allergy and drug intolerance. With GPRO Registry, an assessment of the average number of exceptions were calculated, but the types reported were not available.

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- The measure is specified for EHR and registry.
- All data elements are in defined fields in EHRs.
- For the eMeasure specifications:
  - An eMeasure Feasibility Scorecard was provided. All coding is available in current uses, though ICD-10 & SNOMED-CT codes are not implemented across all EHRs currently. It is not clear how many and what type of EHRs were used for the assessment.
  - Atrioventricular Block and Cardiac Pacer in Situ are listed as exceptions in the eMeasure specification, though they are not included in the registry specifications.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- All data elements are specified using VSAC codes. The Bonnie feasibility scorecard shows this to be feasible.
- At this time the measure is fully specified for EHR and registry, with all the fields being fully defined. Per the requirement, an eMeasure Feasibility Score Card was provided, thus demonstrating acceptable feasibility across multiple EHR systems and sites.

Criterion 4: Usability and Use

4. Usability and 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.

- The measure is currently used in PQRS, Meaningful Use Stage II and the PINNACLE Registry for quality improvement.
- The developer states not unintended consequences have been identified with measure use, and they continuously monitor for applicable mitigation.
- In 2012, the Measure Applications Partnership (MAP) Hospital Workgroup supported the measure stating the measure is strongly tied to outcomes. In 2014, the Clinician Workgroup supported the measure for the Physician Compare and Value-Based Payment Modifier Program stating the measure promotes person- and family-centered care. Promotes alignment across programs, settings, and public- and private-sector efforts. Included in a MAP family of measures. Addresses program goals/requirements. The measure was previously supported by Workgroup for inclusion in Physician Compare and VBPM for clinician group reporting.

Questions for the Committee*:*

- *Is the measure publicly reported?*

- *For maintenance measures – is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- It is currently used in PQRS, MU Stage 2, and the PINNACLE Registry. They mention that CMS may adopt this for Physician Compare public reporting. The developers do not report any known unintended consequences.
- This measure is currently used in PQRS and Meaningful Use Stage II. The measure is also captured and reported via the PINNACLE Registry as a part of QI initiatives as well. TO date, there are been no unintended consequences associated with this measure. The MAP in 2014 reviewed the measure and made recommendations for inclusion in Physician Compare and VBPM

Criterion 5: Related and Competing Measures

- 0070: 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 differ."

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0083

Measure Title: **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: 6/30/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

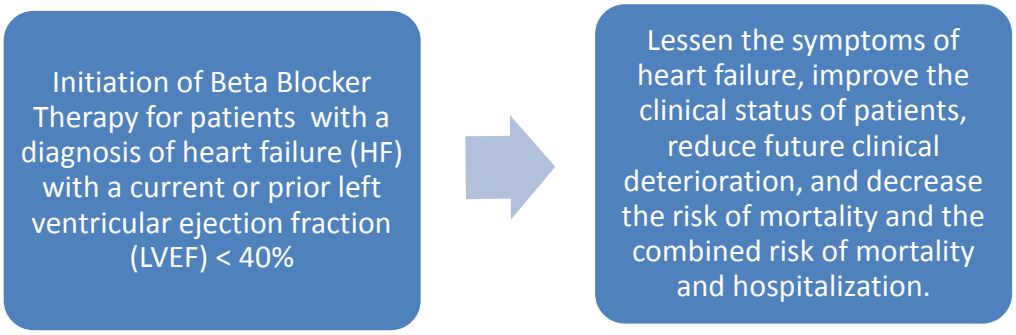
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

http://circ.ahajournals.org/content/128/16/e240

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.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

7.3.2.4. Beta Blockers: Recommendation

Class I

1. Use of 1 of the 3 beta blockers proven to reduce mortality (eg, bisoprolol, carvedilol, and sustainedrelease

metoprolol succinate) is recommended for all patients with current or prior symptoms of HF*r*EF, unless contraindicated, to reduce morbidity and mortality.[346,416–419,448] ***(Level of Evidence: A)***

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

Class I, Level A =

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Table 1. Applying Classification of Recommendation and Level of Evidence

ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT	SIZE OF TREATMENT EFFECT: CLASS I *Benefit >>> Risk* Procedure/Treatment **SHOULD** be performed/ administered	CLASS IIa *Benefit >> Risk* *Additional studies with focused objectives needed* **IT IS REASONABLE** to perform procedure/administer treatment	CLASS IIb *Benefit ≥ Risk* *Additional studies with broad objectives needed; additional registry data would be helpful* Procedure/Treatment **MAY BE CONSIDERED**	CLASS III *No Benefit* or CLASS III *Harm* Procedure/Test / Treatment COR III: No benefit — Not Helpful / No Proven Benefit COR III: Harm — Excess Cost w/o Benefit or Harmful / 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
Suggested phrases for writing recommendations	should is recommended is indicated is useful/effective/beneficial	is reasonable can be useful/effective/beneficial is probably recommended or indicated	may/might be considered may/might be reasonable usefulness/effectiveness is unknown/unclear/uncertain or not well established	COR III: No Benefit: is not recommended; is not indicated; should not be performed/ administered/ other; is not useful/ beneficial/ effective COR III: Harm: potentially harmful; causes harm; associated with excess morbidity/mortality; should not be performed/ administered/ other
Comparative effectiveness phrases†	treatment/strategy A is recommended/indicated in preference to treatment B treatment A should be chosen over treatment B	treatment/strategy A is probably recommended/indicated in preference to treatment B it is reasonable to choose treatment A over treatment B		

A recommendation with Level of Evidence B or C does not imply that the recommendation is weak. Many important clinical questions addressed in the guidelines do not lend themselves to clinical trials. Although randomized trials are unavailable, there may be a very clear clinical consensus that a particular test or therapy is useful or effective.

*Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use.

†For comparative effectiveness recommendations (Class I and IIa; Level of Evidence A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

http://circ.ahajournals.org/content/suppl/2013/06/04/CIR.0b013e31829e8776.DC1/Online_Data_Supplement.pdf

---

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Effectiveness of beta blockers at improving quality of life and clinical status in patients with heart failure; which patients should receive beta blocker therapy; initiation and maintenance of therapy; and risks of treatment with beta blocker therapy.

1a.7.2. Grade assigned for the quality of the quoted evidence <u>with definition</u> of the grade:

Level A-

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

See table in 1a.4.4

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 1989-2009

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

17 Randomized Controlled Trials, 3 comparative studies

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

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 IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

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, beta blockers can reduce the risk of death and the combined risk of death or hospitalization

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

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.

See Online Data Supplement 20 for additional data on beta blockers.
http://circ.ahajournals.org/content/suppl/2013/06/04/CIR.0b013e31829e8776.DC1/Online_Data_Supplement.pdf

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

The articles supporting the Beta Blocker recommendation were from 1989-2009. However, the overall literature search was through Oct, 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 uncertain. They meta-analysed individual-patient data to assess the efficacy of β blockers in patients with heart failure and sinus rhythm compared with atrial fibrillation.

They extracted individual-patient data from ten randomised controlled trials of the comparison of β blockers versus placebo in heart failure. The presence of sinus rhythm or atrial fibrillation was ascertained from the baseline electrocardiograph. The primary outcome was all-cause mortality. Analysis was by intention to treat. Outcome data were meta-analysed with an adjusted Cox proportional hazards regression. The study is registered with Clinicaltrials.gov, number 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 baseline. 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 rhythm and 21% (633 of 3064) in patients with atrial fibrillation. β-blocker therapy led to a significant reduction in all-cause mortality in patients with sinus rhythm (hazard ratio 0·73, 0·67-0·80; p<0·001), but not in patients with atrial fibrillation (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 efficacy for the primary outcome was noted in all subgroups of atrial fibrillation, including age, sex, left ventricular ejection fraction, 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 medications and not regarded as standard therapy to improve prognosis in patients with concomitant heart failure and atrial fibrillation.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form NQF_measure_submission_evidence_HF_BB.docx
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., the benefits or improvements in quality envisioned by use of this measure)* 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 endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.* 2013 PQRS Experience Report*: Data Source: 2013 is the most recent year for which PQRS Experience Report measure data is available. The average performance rates on Heart Failure (HF) – Beta Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD) over the last several years are as follows: 2010: 82.7%

2011: 75.8%
2012: 86.8%
2013: 77.6%
2013 Small Group Practice Exception Rate: 1.04%
Reference: Center for Medicare and Medicaid Services. 2013 Reporting Experience Including Trends. Available: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/PQRS/
*It is important to note that PQRS was a voluntary reporting program, with approximately 51% of eligible professionals participating using any reporting option in 2013, and performance rates may not be nationally representative.

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.**
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.
http://circ.ahajournals.org/content/122/6/585.full
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 endorsement maintenance. 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 subcriterion on improvement (4b.1) 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 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
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

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Severity of illness, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare.**

List citations in 1c.4.

Heart Failure affects over 5.7 million Americans (2.4% in 2008; 2.7% in 2010) and that number is expected to rise consistently over the next 15 years (Heidenreich et al 2011, Mozaffarian 2015), The AHA forecasting study predicts that total costs for heart failure in the 18-44 age group will increase from $1.51 billion to $2.48 billion, while the costs for the 65-79 age group will increase from $11.50 billion to $29.9 billion (Heidenreich et al 2011).

In a 2014 article, Storrow et al writes that heart failure results in nearly 1 million annual hospital stays (Go 2013, Chen 2011), and is the top reason for Medicare hospital readmissions (Jencks 2009, Dharmarajan 2013). The vast majority of patients hospitalized for acute heart failure (AHF) are originally evaluated and managed in the emergency department (ED). Prior data suggest more than 80% of ED patients with AHF are admitted to the hospital and have a median inpatient length of stay (LOS) of approximately 3.4 days. Of the $39.2 billion dollars spent on heart failure care in the United States in 2010, hospital stay was the single largest proportion of this expenditure (AHA 2010, Heidenreich 2011). Among Medicare beneficiaries, hospital stay accounts for more than 50% of all heart failure costs in the last 6 months of life (Blecker 2012). Despite a small decline in the AHF hospital stay rate among Medicare beneficiaries over the last decade (Go 2013, Chen 2011), mortality remains high (Chen 2011, Richardson 2002, Roger 2004) and uneven across

Storrow et al (2014) utilized Nationwide Emergency Department Sample AHF data from 2006 to 2010 to describe admission proportion, hospital length of stay (LOS), and ED charges as a surrogate for resource utilization. Results were compared across U.S. regions, patient insurance status, and hospital characteristics. They concluded that a very high proportion of ED patients with AHF are admitted nationally, with significant variation in disposition and procedural decisions based on region of the country and type of insurance, even after adjusting for potential confounding.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

AHA Writing Group Members , Lloyd-Jones D., Adams R.J., Brown T.M., et al; Heart disease and stroke statisticsd2010 update: a report from the American Heart Association. Circulation. 2010;121:e46-215.
Blecker S., Herbert R., Brancati F.L.; Comorbid diabetes and end-of-life expenditures among Medicare beneficiaries with heart failure. J Card Fail. 2012;18:41-46.
Chen J., Normand S.L., Wang Y., Krumholz H.M.; National and regional trends in heart failure hospital stay and mortality rates for Medicare beneficiaries, 1998–2008. JAMA. 2011;306:1669-1678.
Dharmarajan K., Hsieh A.F., Lin Z., et al; Diagnoses and timing of 30-day readmissions after hospital stay for heart failure, acute myocardial infarction, or pneumonia. JAMA. 2013;309:355-363.
Go A.S., Mozaffarian D., Roger V.L., et al; Heart disease and stroke statistics—2013 update: a report from the American Heart Association. Circulation. 2013;127:e6-e245.
Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, Finkelstein EA, Hong Y, Johnston SC, Khera A, Lloyd-Jones DM, . Nelson SA, Nichol G, Orenstein D, Wilson PWF, Woo YJ on behalf of the American Heart Association Advocacy Coordinating Committee Stroke Council Council on Cardiovascular Radiology and Intervention Council on Clinical Cardiology Council on Epidemiology and Prevention Council on Arteriosclerosis Thrombosis and Vascular Biology Council on Cardiopulmonary Critical Care Perioperative and Resuscitation Council on Cardiovascular Nursing Council on the Kidney in Cardiovascular Disease Council on Cardiovascular Surgery and Anesthesia, and Interdisciplinary Council on Quality of Care and Outcomes Research. Forecasting the Future of Cardiovascular Disease in the United States: A Policy Statement From the American Heart Association. Circulation, March 1, 2011 vol. 123 no. 8 933-944. http://circ.ahajournals.org/content/123/8/933/T1.expansion.html

Jencks S.F., Williams M.V., Coleman E.A.; Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009;360:1418-1428.
Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation 2015;131:e29–322. doi:10.1161/CIR.0000000000000152
Richardson L.D., Asplin B.R., Lowe R.A.; Emergency department crowding as a health policy issue: past development, future directions. Ann Emerg Med. 2002;40:388-393.
Roger V.L., Weston S.A., Redfield M.M., et al; Trends in heart failure incidence and survival in a community-based population. JAMA. 2004;292:344-350.

Storrow AB, MD; Jenkins CA; Self WH; Alexander PT; Barrett TW; Han JH; McNaughton CD; Heavrin BS; Gheorghiade M; Collins SP. The Burden of Acute Heart Failure on U.S. Emergency Departments. JCHF. 2014;2(3):269-277. doi:10.1016/j.jchf.2014.01.006

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable. Not a PRO-PM.

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

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.)*
eCQM Library webpage at: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/eCQM_Library.html
Value set details at VSAC webpage: 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:** EP_CMS144v4_NQF0083_HF_BB-635712712817825444.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:** 0083_AMAPCPI_HF-BB_ValueSets_June2015-635712735683880063.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Supporting guidelines and coding value sets included in the measure are reviewed on an annual basis. This annual review resulted in very limited changes to adhere to current eCQM industry standards and preserve the original measure intent.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who were prescribed* beta-blocker therapy** either within a 12 month period when seen in the outpatient setting or at hospital discharge

*Prescribed may include:
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
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

**Beta-blocker therapy should include bisoprolol, carvedilol, or sustained release metoprolol succinate. (see technical specifications for additional information on medications)

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
At least once during the 12 consecutive month measurement period when seen in the outpatient setting OR at each hospital discharge during the 12 consecutive month measurement period if seen in the inpatient setting

S.6. **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, 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.
For EHR:
HQMF eMeasure developed and is included in this submission.

For Registry:
Definitions:
Prescribed – Outpatient Setting - May include 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: May include 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.
Beta-blocker Therapy - For patients with prior LVEF < 40%, beta-blocker therapy should include bisoprolol, carvedilol, or sustained release metoprolol succinate.

Report Quality Data Code, G8450: Beta-blocker therapy prescribed

S.7. **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%

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

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
For EHR:
HQMF eMeasure developed and is included in this submission.

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

DENOMINATOR NOTES:
To meet this measure, it must be reported for all heart failure patients a minimum of once during the measurement period when seen in the outpatient setting AND reported at each hospital discharge during the measurement period.

The requirement of "Count >=2 of Encounter, Performed" is to establish that the eligible professional has an existing relationship with the patient.

For Registry:
Option 1, Outpatient Setting:
Patients aged >=18 years
AND
Diagnosis for heart failure (ICD-9-CM) [for use 1/1/2015-9/30/2015]: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13,

404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9
Diagnosis for heart failure (ICD-10-CM) [for use 10/01/2015-12/31/2015]: I11.0, I13.0, I13.2, I50.1, I50.20, I50.21, I50.22, I50.23, I50.30, I50.31, I50.32, I50.33, I50.40, I50.41, I50.42, I50.43, I50.9
AND
Patient encounter(s) during the reporting period (CPT): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99304, 99305, 99306, 99307, 99308, 99309, 99310, 99324, 99325, 99326, 99327, 99328, 99334, 99335, 99336, 99337, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350
AND
Two Denominator Eligible Visits
AND
Left ventricular ejection fraction (LVEF) < 40% or documentation of moderately or severely depressed left ventricular systolic function: G8923

Option 2, Inpatient Setting:
Patients aged >= 18 years
AND
Diagnosis for heart failure (ICD-9-CM) [for use 1/1/2015-9/30/2015]: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9
Diagnosis for heart failure (ICD-10-CM) [for use 10/01/2015-12/31/2015]: I11.0, I13.0, I13.2, I50.1, I50.20, I50.21, I50.22, I50.23, I50.30, I50.31, I50.32, I50.33, I50.40, I50.41, I50.42, I50.43, I50.9
AND
Patient encounter during reporting period (CPT): 99238, 99239
AND
Left ventricular ejection fraction (LVEF) < 40% or documentation of moderately or severely depressed left ventricular systolic function: 3021F

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
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)
Documentation of patient reason(s) for not prescribing beta-blocker therapy
Documentation of system reason(s) for not prescribing beta-blocker therapy

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
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. This measure was developed using the PCPI exception methodology which 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, exceptions may 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), Documentation of patient reason(s) for not prescribing beta-blocker therapy, or Documentation of system reason(s) 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 eSpecifications. 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.

Additional details by data source are as follows:

For EHR:
HQMF eMeasure developed and is included in this submission.

For Registry:
Report Quality Data Code G8451: Beta-Blocker Therapy for LVEF < 40% not prescribed for reasons documented by the clinician (eg, low blood pressure, fluid overload, asthma, patients recently treated with an intravenous positive inotropic agent, allergy, intolerance, other medical reasons, patient declined, other patient reasons, other reasons attributable to the healthcare system)

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 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.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
n/a

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)

S.16. Type of score:
Rate/proportion
If other:

S.17. 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.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*
To calculate performance rates:
1. Find the patients who meet the initial population (ie, 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 (ie, 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 (ie, 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: 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); Documentation of patient reason(s) for not prescribing beta-blocker therapy; Documentation of system reason(s) 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 (ie, 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.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable. The measure is not based on a sample.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. The measure is not based on a survey.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Patient eligibility is determined by a set of defined criteria relevant to a particular measure. If data required to determine patient eligibility are missing, those patients/cases would be ineligible for inclusion in the denominator and therefore the patient/case would be deleted.

If data required to determine if a denominator eligible patient qualifies for the numerator (or has a valid exclusion/exception) are missing, this case would represent a quality failure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Home Health, Hospital/Acute Care Facility, Other, Post Acute/Long Term Care Facility : Long Term Acute Care Hospital, Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility
If other: Domiciliary

S.28. **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.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_0083_Heart_Failure_-HF-_-_Beta_Blocker_Therapy_for_LVSD_Testing_Attachment.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: Heart Failure – Beta Blocker Therapy for Left Ventricular Systolic Dysfunction
Date of Submission: Click here to enter a date

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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.

2b2. 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.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

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.

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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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.23*)**	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/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*).

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 have 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)

The data source is the Centers for Medicare & Medicaid Services (CMS) PQRS GPRO database.

Data 4 (EHR – Exceptions Analysis)

The data source is EHR data.

1.3. What are the dates of the data used in 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 3 (GPRO Registry)

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.

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.26***)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

Data 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 3 (GPRO Registry)

The total number of physicians reporting on this measure is 1,748. Of those, 684 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, 39.1 percent of physicians are included in the analysis, and the average number of quality reporting events is 33.9 for a total of 23,175 events. The range of quality reporting events for 684 physicians included is from 326 to 10. The average number of quality reporting events for the remaining 60.9 percent of physicians who aren't included is 3.2.

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.

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

Data 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 3 (GPRO Registry)

There were 23,175 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.

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

Data 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.

Data 3 (GPRO Registry)

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.

1.8. What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

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.

Data 3 (GPRO Registry)

This was not captured as part of the testing.

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*)

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.

2a2.3. For each level 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*)

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.

Data 3 (GPRO Registry)

For this measure, the reliability at the minimum level of quality reporting events (10) was 0.86. The average number of quality reporting events for physicians included is 33.9. The reliability at the average number of quality reporting events was 0.96.

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?*)

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.

Data 3 (GPRO Registry)

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

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. For each level 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)*

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

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

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

2b2.4. 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?*)

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)

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

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.

2b3.2. 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*)

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 3 (GPRO Registry)

Amongst the 684 physicians with the minimum (10) number of quality reporting events, there were a total of 1,203 exceptions reported. The average number of exceptions per physician in this sample is 1.8. 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.

2b3.3. 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*)

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.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities)

Not applicable

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. 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)*

Not applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

Not applicable

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. 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*)

Not applicable

***2b4.11. 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*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

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.

Data 3 (GPRO Registry)

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

2b5.2. 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*)

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).

Data 3 (GPRO Registry)

Based on the sample of 684 included physicians, the mean performance rate is 0.70, the median performance rate is 0.93 and the mode is 1.00. The standard deviation is 0.37. The range of the performance rate is 1.00, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.58 (0.42 - 1.00).

2b5.3. 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?*)

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.

Data 3 (GPRO Registry)

The range of performance from 0.00 to 1.00 suggests there's clinically meaningful variation across physicians' performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

This test was not performed for this measure.

2b6.2. 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*)

This test was not performed for this measure.

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

This test was not performed for this measure.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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)

Data are not available to complete this testing.

2b7.2. 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)

Data are not available to complete this testing.

2b7.3. 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)

Data are not available to complete this testing.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
3a. Byproduct of Care Processes

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

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

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.**
Attachment **Attachment:** NQF_0083_Feasibility_Scorecard_Bonnie_Output_Screen_Shots_Revised.pdf

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM 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.

Planned	Current Use (for current use provide URL)
	Public Reporting Physician Quality Rating System http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/pqrs/index.html Payment Program Meaningful Use Stage II http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Stage_2.html Quality Improvement with Benchmarking (external benchmarking to multiple organizations) PINNACLE Registry http://cvquality.acc.org/en/NCDR-Home/Registries/Outpatient-Registries.aspx

4a.1. For each CURRENT use, checked above, provide:

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

1) Physician Quality Reporting System (PQRS)-Sponsored by the Centers for Medicare and Medicaid Services (CMS)
Purpose: PQRS is a national reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs). The program provides an incentive payment to practices with EPs (identified on claims by their individual National Provider Identifier [NPI] and Tax Identification Number [TIN]). Eps satisfactorily report data on quality measures for covered Physician Fee Schedule (PFS) services furnished to Medicare Part B Fee-for-Service (FFS) beneficiaries (including Railroad Retirement Board and Medicare Secondary Payer). Beginning in 2015, the program also applies a payment adjustment to EPs who do not satisfactorily report data on quality measures for covered professional services. Source: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html It is our understanding that CMS is also planning to move towards publicly
reporting physician data via Physician Compare.

2) Meaningful Use Stage 2 (EHR Incentive Program) – Sponsored by the Centers for Medicare and Medicaid Services (CMS)
The Medicare and Medicaid EHR Incentive Programs provide incentive payments to eligible professionals, eligible hospitals, and critical access hospitals (CAHs) as they adopt, implement, upgrade or demonstrate meaningful use of certified EHR technology.

These professionals are eligible for incentive payments for the "meaningful use" of certified EHR technology, if all program requirements are met, including successful implementation and reporting of program measures, which include this measure, to demonstrate meaningful use of EHR technology.

3) PINNACLE Registry (URL: http://cvquality.acc.org/en/NCDR-Home/Registries/Outpatient-Registries.aspx)
The PINNACLE Registry® is cardiology´s largest outpatient quality improvement registry, capturing data on coronary artery disease, hypertension, heart failure and atrial fibrillation. The PINNACLE Registry® continues to grow rapidly, with more than 2400 providers representing almost 800 unique office locations across the U.S submitting data to the registry as of the fourth quarter of 2013. As of the fourth quarter of 2013, the registry has more than 13 million patient encounter records. PINNACLE assists practices in understanding and improving care through the production and distribution of quarterly performance reports. These reports, covering all valid patient encounters, detail adherence to 28 cardiovascular clinical measures at the physician, location, and practice levels across coronary artery disease, hypertension, heart failure and atrial fibrillation. All jointly developed ACC/AHA/PCPI performance measures for these topics are reported by the registry.

4a.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 new insurance marketplace.

4a.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. 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.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

2013 PQRS Experience Report*:
Data Source: 2013 is the most recent year for which PQRS Experience Report measure data is available. The average performance rates on Heart Failure (HF) – Beta Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD) over the last several years are as follows:

2010: 82.7%
2011: 75.8%
2012: 86.8%
2013: 77.6%

2013 Small Group Practice Exception Rate: 1.04%
Reference: Center for Medicare and Medicaid Services. 2013 Reporting Experience Including Trends. Available: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/PQRS/
*It is important to note that PQRS was a voluntary reporting program, with approximately 51% of eligible professionals participating using any reporting option in 2013, and performance rates may not be nationally representative.

4b.2. **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.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We are not aware of any unintended consequences at this time, but we take unintended consequences very seriously and therefore continuously monitor to identify actions that can be taken to mitigate them.

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%)
0071 : Persistence of Beta-Blocker Treatment After a Heart Attack

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

5a. Harmonization

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 completely harmonized?
No

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

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):** AMA-PCPI
Co.2 **Point of Contact:** Caryn, Davidson, caryn.davidson@ama-assn.org, 312-464-4465-
Co.3 **Measure Developer if different from Measure Steward:** AMA-PCPI
Co.4 **Point of Contact:** Caryn, Davidson, pcpimeasures@ama-assn.org

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
PCPI and ACC/AHA measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study are invited to participate as equal contributors to the measure development process. In addition, the PCPI and ACC/AHA strive to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

Work Group members:
Craig T. Beam, CRE (patient representative)
Ileana L. Piña, MD, FACC (cardiology, heart failure)
Kathleen Blake, MD (cardiac electrophysiology)
Paul D. Rockswold, MD, MPH (family medicine)
Donald E. Casey, Jr., MD, MPH, MBA, FACP, FAHA (internal medicine)
Lawrence B. Sadwin (patient representative)
Sarah J. Goodlin, MD (geriatrics, palliative medicine)
Joanna D. Sikkema, MSN, ANP-BC, FAHA (cardiology)
Kathleen L. Grady, PhD, APN, FAAN, FAHA (cardiac surgery)
Carrie A. Sincak, PharmD, BCPS (pharmacy)
Randal F. Hundley, MD, FACC (cardiology, health plan representative)
John Spertus, MD, MPH (cardiology)
Mariell Jessup, MD, FACC, FAHA, FESC (cardiology, heart failure)
Patrick J. Torcson, MD, FACP, MMM (hospital medicine)
Thomas E. Lynn, MD (family medicine, measure implementation)
Elizabeth Torres, MD (internal medicine)
Frederick A. Masoudi, MD, MSPH (cardiology)
Mark V. Williams, MD, FHM (hospital medicine)
David Nilasena MD, MSPH, MS (general preventive medicine, public health, measure implementation)
John B Wong, MD (internal medicine)
American College of Cardiology Foundation
Charlene L. May
Melanie Shahriary, RN, BSN

American Heart Association
Cheryl Perkins, MD, RPh
Mark D. Stewart, MPH
Gayle Whitman, PhD, RN, FAHA, FAAN

American College of Cardiology
Foundation/American Heart Association
Jensen S. Chiu, MHA
National Committee for Quality Assurance (NCQA) Liaison
Manasi Tirodkar, PhD, MS

The Joint Commission Liaison
Millie J. Perich, MS, RN
American Medical Association
Mark Antman, DDS, MBA
Heidi Bossley, MSN, MBA
Christopher Carlucci, MBA
Kerri Fei, MSN, RN
JoeAnn Jackson, MJ
Kendra Hanley, MS
Karen Kmetik, PhD
Pamela O'Neil, MPH
Samantha Tierney, MPH
Temaka Williams, MPH, MBA
Greg Wozniak, 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:** 12, 2014

Ad.4 **What is your frequency for review/update of this measure?** Coding/Specifications updates occur annually.

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2015

Ad.7 **Disclaimers:** Physician Performance Measures (Measures) and related data specifications have been developed by the American Medical Association (AMA) - convened Physician Consortium for Performance Improvement(R) (PCPI[R]), American College of Cardiology (ACC) and American Heart Association (AHA). These Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Measures require a license agreement between the user and the AMA, (on behalf of the PCPI), ACC or AHA. Neither the AMA, ACC, AHA, PCPI nor its members shall be responsible for any use of the Measures.

THE MEASURES 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 disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT[R]) or other coding contained in the specifications.

Ad.8 **Additional Information/Comments:**

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 #: 0081
Measure Title: Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)
Measure Steward: AMA-PCPI
Brief Description of Measure: Patients who were prescribed* ACE inhibitor or ARB therapy either within a 12 month period when seen in the outpatient setting or at hospital discharge
Developer Rationale: In the absence of contraindications, ACE inhibitors or ARBs are recommended for all patients with symptoms of heart failure and reduced left ventricular systolic function. ACE inhibitors remain the first choice for inhibition of the renin-angiotensin system in chronic heart failure, but ARBs can now be considered a reasonable alternative. Both pharmacologic agents have been shown to decrease the risk of death and hospitalization. Additional benefits of ACE inhibitors include the alleviation of symptoms and the improvement of clinical status and overall sense of well-being of patients with heart failure.

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.

Numerator Statement: Patients who were prescribed* ACE inhibitor or ARB therapy either within a 12 month period when seen in the outpatient setting or at hospital discharge

*Prescribed may include:
Outpatient setting: prescription given to the patient for ACE inhibitor or ARB therapy at one or more visits in the measurement period OR patient already taking ACE inhibitor or ARB therapy as documented in current medication list
Inpatient setting: prescription given to the patient for ACE inhibitor or ARB therapy at discharge OR ACE inhibitor or ARB therapy to be continued after discharge as documented in the discharge medication list
Denominator Statement: All patients aged 18 years and older with a diagnosis of heart failure with a current or prior LVEF < 40%
Denominator Exclusions: Documentation of medical reason(s) for not prescribing ACE inhibitor or ARB therapy (eg, hypotensive patients who are at immediate risk of cardiogenic shock, hospitalized patients who have experienced marked azotemia, allergy, intolerance, other medical reasons)
Documentation of patient reason(s) for not prescribing ACE inhibitor or ARB therapy (eg, patient declined, other patient reasons)
Documentation of system reason(s) for not prescribing ACE inhibitor or ARB therapy (eg, other system reasons)

Measure Type:
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry
Level of Analysis: Clinician : Group/Practice, Clinician : Individual
Is this an eMeasure? ☒ **Yes** ☐ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☒ **Yes** ☐ **No**

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: 8/10/09 **Most Recent Endorsement Date:** 1/18/12

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

Public and Member Comments

- The excessive patient, system, and medical exclusions in this measure should be revisited so that they all meet the following criteria: evidence-based, highly specific, and explicitly defined.
- Obtaining data to calculate these measures could be challenging for certain end users. Prescription of ACE inhibitor or ARB therapy is occurring at the time of hospital discharge, however to collect the data for individual clinicians would be very labor intensive. Measuring this at both levels may lead to duplication of medications and increase medication errors.
- Suggest limiting to specific drugs that are FDA approved for use in HF/LVSD: ARBs: candesartan (has a mortality claim) and valsartan.
- An ARB should be used when available for black patients as ACEI in black patients cause more angioedema.

Developer response:

- These measures have been tested and found to be generally feasible in EHR, paper, and claims data sources. This is a clinician-level measure for the outpatient setting.
- As specified, this measure applies to patients with CAD and LVSD OR patients with CAD and diabetes. The list of medications/drug names included in the measure specifications is based on clinical guidelines and other evidence. The specified drugs were selected based on the strength of evidence for their clinical effectiveness. Available data suggests that there are no differences among available ACEIs and ARBs in their effects on symptoms or survival.
- This measure is intended to encourage ACEI or ARB therapy in the treatment of patients with HF and LVSD. The specific type of ACEI or ARB prescribed is at the discretion of the clinician and should be specific type of ACEI or ARB prescribed is at the discretion of the clinician and should be specific to the needs of the individual patient.

Steering Committee: Reviewed comments and developer's responses. No change in recommendations.

IF this measure is paired/grouped, list the paired measure NQF#/title: 0083: Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction

IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Measures #0083 (Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction) and #0081 address related aspects of care for effective treatment for patients with heart failure and should be measured concurrently. Both ACE inhibitors and beta-blockers have been shown to reduce mortality and hospitalizations and improve a patient's clinical status. ARBs can be considered a reasonable alternative for ACE inhibitors. Combined treatment with these agents produces additive benefits and is required for optimal management of heart failure.14 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.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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.

The developer provides the following evidence for this process measure:

- This clinician-level process *registry measure* and *eMeasure* calculates the percentage of patients aged 18 years and older with a diagnosis of heart failure with a current or prior LVEF < 40% who were prescribed an ACE Inhibitor or ARB therapy either within a 12 month period when seen in the outpatient setting or at hospital discharge.
- The developer provides the 2013 ACCF/AHA Guideline for the management of heart failure (Class I, Level A) with one recommendation for ACE Inhibitors (Class I) and four recommendations for ARBs in patients who are: ACE intolerant (Class I), already taking ARBs for other indications (Class IIa), addition of an ARB (Class IIb) and

recommendation to avoid the combined use of ACE, ARB, and aldosterone antagonist (Class III).

- The articles supporting the ACE/ARB recommendations were from 1993-2012 but the developers report that the overall literature search was through October 2011 with select articles included through April 2013. The developer states that additional articles found from 2014 and 2015 would not change the ACE/ARB recommendations.
- QQC - 2 meta-analyses, 10 randomized controlled trials, 3 comparative studies, and 1 review paper supporting the ACE/ARB recommendations.
- The developer provides a diagram of how the initiation of ACE Inhibitor therapy for patients with heart failure or LVEF < 40% is linked to patient outcomes.

Questions for the Committee:

- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- The developers provide the average performance rates from the PQRS Experience Report from 2010-2013 and data from the literature that shows an ACEI/ARB was prescribed for 11,165 (79.8%) of 13,987 eligible patients. Additional data (std dev, min, max, interquartile range, scores by decile, number of measured entities) that is required per NQF policy for maintenance is not provided. The developer should clarify if the PINNACLE data collection tool includes race, and other SES factors, such as insurance status.
- The developers report that disparities data from the federal reporting programs using this measure have not yet been made available for them to analyze and report, though they encourage the results of this measure to be stratified by race, ethnicity, sex, and payer.
- The developers provide data from a 2011 study that suggests that there are some racial and ethnic disparities in the receipt of pharmacological therapy for CHF among TRICARE beneficiaries.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - Broad and convincing
   - Yes the evidence supports the use of ACE-I or ARB in LVSD based on clinical trial evidence and ACC/AHA guideline recommendations
   - Yes, the process of care is related to the desired outcome, but actual filling of medication would be a better marker than just an rx for the ACE-I or ARB
   - No, there are not exceptions

- The measure is supported by the most recent ACC/AHA HF guidelines from 2013.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***
- Solid consistent evidence
- The guideline supporting evidence directly supports the process being evaluated as it relates to ACE-I or ARB prescribed in individuals with LVSD.

1b. ***Committee's Comments on Performance Gap:***
- Disparities are well demonstrated
- Yes, there is a performance gap as the reported Rx rate of 79.8% is low when looking at the context that it's a Class IA recommendation in guidelines and has been for many years,
- Ethnicity data would be useful to evaluate if black patients are less likely to get prescribed and ACEI/ARB secondary to a higher adverse event rate compared to other races
- Disparity data not available but would be useful

1c. ***Committee's Comments on Composite Performance Measure:***
- The overall, quality construct is logical and aligns with the measure.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure's data source is an EHR and/or registry. ICD 9, ICD 10 and CPT codes provided for the numerator and denominator for registry and inpatient setting. ICD-10 conversion methodology is not discussed. In both sets of specifications (registry and eMeasure), the logic is unambiguous. Higher scores equal better quality.
- In the outpatient setting, "prescribed" may include an ACEI/ARB prescription given to the patient at one or more visits in the measurement period or ACEI/ARB listed as current medication. In the inpatient setting, "prescribed" may include prescription given at discharge or ACEI/ARB listed in discharge medication list.
- In the outpatient setting, 2 or more encounters are required to establish the eligible professional has an existing relationship with the patient.
- The measure is intended for use in an office visit, outpatient consultation, nursing facility, long-term care residential facility, home health and provider interaction during the measurement period.
- For the registry specification, the denominator patient, medical & system exclusions/exceptions are reported using separate CPT-II codes for each of the 3 type of exclusions/exceptions, though the paired measure 0083 uses a single HCPCS code G8541 to report all 3 types of exclusions/exceptions.
- Denominator exclusions/exceptions include broad definitions for medical reason, patient reason and system reason for not prescribing ACEI/ARB; examples of exceptions provided using CPT-II codes. The developer states that exceptions should only be considered when the numerator activity was not performed, that they are not uniformly relevant across measures, and that there must be a clear rationale to permit an exception for a medical, patient, or system reason. In the provided value sets, broadly defined and inappropriate patient, medical and system reason denominator exclusions include but are not limited to: medical reason: 216952002 failure in dosage (event), patient reason: 224187001 variable income (finding) & 266966009 family illness (situation), system reason:266756008 medical care unavailable (situation).
- Missing numerator data represent a quality failure.
- The calculation algorithm is included.
- The measure is not risk adjusted and SDS variables were not captured for the measure. The developer encourages users to provide collect data and stratify results by race, ethnicity, administrative sex, and payer consistent with CMS' Measures Management System Blueprint and recent national recommendations put forth by the IOM and NQF.

- For the eMeasure, the developer should clarify the following:
    - Pregnancy & Renal Failure due to ACE Inhibitor are included in value sets as exceptions, and not listed in the eMeasure specification, and they are not included in the Registry specification. Should these be exclusions?
    - Atrioventricular Block and Cardiac Pacer in Situ are listed as exceptions in the eMeasure specification, though they are not included in the registry specifications. Should these be exclusions?
    - "Provider interactions" are listed as encounters in the value set spreadsheet and in the eMeasure specifications that include both face-to-face visits and non-face-to-face communications. The developer is encouraged to provide reasoning for inclusion, and clarify if all provider interactions are included in the denominator definition for a patient encounter.
    - "Communication: From Patient to Provider: Patient Reason for ACE Inhibitor or ARB Decline" is listed as an exception and is not included in the Registry specification. Please describe intent of this exception.
    - In the Initial Patient Population (IPP), "Encounter, Performed: Patient Provider Interaction" is listed, though it is not in the Data Criteria (QDM Variables).
    - "Encounter, Performed: Face-to-Face Interaction" is listed in the Data Criteria (QDM Variables), though not in the IPP.
    - Medication initial, maximum & mean does are provided in the header information. The developer should clarify if this is for information purposes, or part of the eMeasure.
- All eMeasure specifications and values sets meet current NQF eMeasure technical requirements and are provided on Sharepoint for SC review.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

GPRO Registry Testing

- The developer tested reliability at the performance measure score level, using a beta-binomial model in a signal-to-noise analysis, to differentiate the true differences between measured entities (the signal) from random measurement error (the noise). A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in performance (for this measure, between providers). A value of 0.7 is often regarded as a minimum acceptable reliability value.
- A sample 1,244 (33.4%) of 3,728 physicians reported the measure with results showing the average number of quality reporting events for physicians included is 31.5 for 39,242 events with a reliability of 0.94 (high reliability). For the program required minimum of 10 quality reporting events, with a reliability of 0.83 (high reliability).
- A missing data assessment was not performed. The developer states data missing from denominator excludes the patient from the measure, while missing numerator data counts as a measure "fail".

eMeasure Testing

- Critical data element (Validity against the Gold Standard) was conducted. Per NQF criteria, if empirical validity testing was performed of patient-level data, the rating from validity testing of patient-level data elements should be used. The developer provides simple agreement results for critical data element validity. Comparison of the values for several data elements (electronic extracted vs. data abstracted) using EHR data was conducted and could satisfy the data element reliability criterion (Algorithm box 3; validity testing results described in 2b.2 below). Results were provided for most, but not all, critical data elements. Percent agreement statistics were presented; however, percent agreement does not adjust for agreement due to chance, and should not be used alone to demonstrate reliability.

- Ideally, implementation of an eMeasure can be considered an automated process, and therefore the calculations will be consistent. The submitted eMeasure specification follows industry standards to represent the measure electronically which should enable automated data extraction and measure score calculation.
- In addition to critical data element testing, the developer submitted pre-testing from the Measure Authoring Tool within the Bonnie Output that tests eMeasure logic. The measure logic successfully validated through the Bonnie Output.

- For both specifications, a missing data assessment was not performed. The developer states data missing from denominator excludes the patient from the measure, while missing numerator data counts as a measure "fail".
- Patient-level socio-demographic (SDS) variables were not captured as part of the testing.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The clinical practice guidelines supporting this measure recommend the use of ACEI/ARBs in patients with HF.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

GPRO Registry Testing

- The developer provided face validity results using 12 members of the PCPI Measure Advisory Committee (MAC) with a mean rating of 4.33 and 100% of respondents either agree or strongly agree the measure is able to distinguish good and poor quality using a 1-5 Likert, highest score was 5. The MAC is independent of the measure developer experts.

eMeasure Testing

- The developer provided critical data element (Validity against the Gold Standard) was conducted. Per NQF criteria, if empirical validity testing was performed of patient-level data, this rating will also be used for reliability testing. Data elements included patient age, visit, problem list, or medical history of CAD, those who met the denominator population.
- The developers provide simple agreement for the 254 patients sampled via automated EHR review. Of the sample, 217 patients (85.4%) were detected for numerator criteria, 23 patients detected for exclusion criteria totaling 93.9% calculation detection to the gold standard (the EHR) via automated review 6 patients were further added and 2 were false inclusions, totaling 96.1% agreement. Additional statistical testing was not provided.
- Data element validity testing was conducting by comparing, for several data elements, the values obtained from electronic extraction from 1 EHR to those obtained from those abstracted from the EHR by an abstractor. Simple agreement was provided for most, but not all, critical data elements. However, simple agreement does not adjust for agreement due to chance and thus should not be used alone to demonstrate validity; sensitivity/specificity statistics are preferred for demonstrating data element validity. Percentage agreement values were relatively high for most data elements considered. It appears that only one abstractor was utilized, which is acceptable for testing validity against the gold standard in an EHR.
- In addition to critical data element testing, the developer submitted pre-testing from the Measure Authoring Tool within the Bonnie Output that also tests eMeasure performance calculation. This testing does use "live" EHR

patients, though NQF currently accepts Bonnie Output pre-testing when EHR testing was not provided. Results in the 46 "pre-test" patients demonstrated 100% agreement for identifying both expected and actual initial patient population, denominator, denominator exclusions, numerator, and denominator exceptions. Testing characteristics are provided for the 46 "pre-test" patients, with 100% of the of data elements concepts included in the Initial Patient Population (IPP), Denominator, Numerator and Denominator Exceptions, with 100% of possible data elements covered in the pre-test sample.

Questions for the Committee:

- ○ *Is the test sample adequate to generalize for widespread implementation?*
- ○ *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- ○ *Do you agree that the score from this measure as specified is an indicator of quality?*
- ○ *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions: Analysis of exclusions found:

- GPRO Registry: Of the 1244 physicians with the minimum (10) number of quality reporting events, there were a total of 8,056 exceptions reported. The average number of exceptions per physician in this sample is 6.5. The overall exception rate is 17.03%. The types of exceptions reported are not available from the GRPO Registry.
- EHR: Measure exceptions were validated 95.32% of the time. Review of the 127 exceptions revealed 99.5% of exceptions were medical reasons for not prescribing ACE inhibitor or ARB therapy. Medical reason exceptions consisted of clinical contraindications, drug allergy and drug intolerance.
- Broad exceptions (medical, patient & system) reasons are included within the provided data sets, and include non-relevant codes.
- For the eMeasure specifications:
  - Pregnancy & Renal Failure due to ACE Inhibitor are included in value sets as exceptions, and not listed in the eMeasure specification, and they are not included in the Registry specification.
  - Atrioventricular Block and Cardiac Pacer in Situ are listed as exceptions in the eMeasure specification, though they are not included in the registry specifications.
- Broad exceptions (medical, patient & system) reasons are included within the provided data sets, and include non-relevant codes.

Questions for the Committee:

- ○ *Are the exclusions consistent with the evidence?*
- ○ *Are any patients or patient groups inappropriately excluded from the measure?*
- ○ *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This process measure is not risk adjusted.

2b5. Meaningful difference:

- GPRO Registry: Based on the sample of 1,244 included physicians, the mean performance rate is 0.80, the median performance rate is 0.94 and the mode is 1.00. The standard deviation is 0.29. The range of the performance rate is 1.00, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.29 (0.71 - 1.00).
- Meaningful difference data for the eMeasure was not provided by the developer, as the eMeasure testing data used fictitious "pre-test" patients, though higher performance and detection rates were noted in eMeasure validity testing with EHR & manual reviews, than in EHR-only reviews.

Question for the Committee:

- ○ *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- NQF criteria require a comparability assessment of data sources/methods, such as with multiple specifications for the same measure. This was not provided by the developer, as the eMeasure testing data used fictitious "pre-test" patients.
- For validity testing of the eMeasure, developers did find higher performance & detection rates for the records abstracted with EHR & manual reviews, than EHR-only reviews.

2b7. Missing Data

- Missing data represent a quality failure.
- The developer reports "Data are not available to complete this testing."

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- The problem with this measure is that it compares apples and oranges
  - the numerator specifies use of ACEI/ARB in a 12 month period
  - the denominator specifies any EF (ever) of <40%
- I think this should be remedied so that EF is also considered within the same period of time"
- Need to clarify exclusions (pregnancy, renal failure, AV block)
- The logic and calculation algorithm is clear
- It is likely this measure can be implemented in a consistent fashion.

2a2.: ***Committee's Comments on Reliability-Testing:***

- It is reliable
- High reliability score (0.94)

2b1.: ***Committee's Comments on Validity-Specifications***:

- Same answer as 2a1
- Test sample is of great enough size being 1244/3728
- Yes, valid recommendation per ACC/AHA 2013 guidelines

2b2.: ***Committee's Comments on Validity-Testing:***

- It is valid
- Is the test sample adequate to generalize for widespread implementation? Yes, 254 patients
- Do the results demonstrate sufficient validity so that conclusions about quality can be made? Yes, 96.1% agreement.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- See 2a1
- I think the measure needs to be bit refined as numerator and denominator should be from the same period, rather using denominator definition as having at any point in the past LVEF <40%
- Are the exclusions consistent with the evidence? Yes, but need to clarify the eMeasure specifications
- Are any patients or patient groups inappropriately excluded from the measure? NO
- Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? 17.3% Rate of exclusion

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- For both Registry & eMeasures, data elements are collected by and used by healthcare personnel during provisions of care (BP, lab values), codes by someone other than the person obtaining the information (billing), and abstracted by someone other personnel (quality staff). All data elements are in electronic fields.
- The measure is specified for Clinician : Group/Practice, Clinician : Individual use.
- In the eMeasure Feasibility Scorecard, the developer states all data elements score "3" on a scale of 1 – 3 (3 the highest) for current and future use. The developer should clarify if this includes ICD-10, SNOMED-CT & RxNorm codes for the eMeasure for all data elements. The EHR product(s) & number of EHRs used in the eMeasure Feasibility Scorecard is not reported. NQF requires a Feasibility Scorecard from more than one EHR.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- It is feasible.
- Are the required data elements routinely generated and used during care delivery? Yes, collected in route practice
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources? Yes, in route practice
- Is the data collection strategy ready to be put into operational use? Yes, based on the description and feasibility assessment
- If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites? Need a feasibility score from more than one EHR

Criterion 4: Usability and Use

4. Usability and 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.

- The measure is currently used in PQRS, Meaningful Use Stage II and the PINNACLE Registry for quality improvement.
- The developer states not unintended consequences have been identified with measure use, and they continuously monitor for applicable mitigation.
- In 2014, the Measure Applications Partnership (MAP) Clinician Workgroup supported the measure for the Physician Compare and Value-Based Payment Modifier Program stating the measure promotes person- and family-centered care. Promotes alignment across programs, settings, and public- and private-sector efforts. Provides consideration for healthcare disparities and cultural competency. Included in a MAP family of measures. The measure was previously supported by Workgroup for inclusion in Physician Compare and VBPM for clinician group reporting.

Questions for the Committee (as appropriate) *:*

- *Is the measure publicly reported?*
- *For maintenance measures – is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments
Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- It is being used successfully
- Is the measure publicly reported? Yes, from multiple registries
- How can the performance results be used to further the goal of high-quality, efficient healthcare? Could be used to compare institutions
- Do the benefits of the measure outweigh any potential unintended consequences? Yes, as recommendation is a class 1A recommendation per ACC/AHA 2013 HF guidelines

Criterion 5: Related and Competing Measures

- None listed

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Click here to enter measure title

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

Date of Submission: Click here to enter a date

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

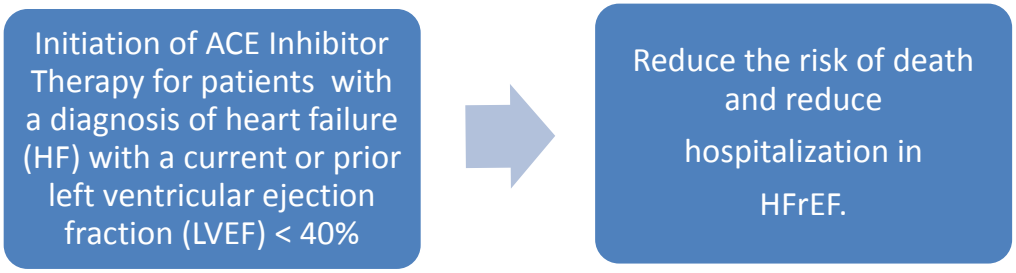
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

http://circ.ahajournals.org/content/128/16/e240

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.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

7.3.2.2. ACE Inhibitors: Recommendation

Class I

1. **ACE inhibitors are recommended in patients with HF*r*EF and current or prior symptoms, unless contraindicated, to reduce morbidity and mortality. 343,412–414 *(Level of Evidence: A)***

7.3.2.3. ARBs: Recommendations

Class I

1. **ARBs are recommended in patients with HFrEF with current or prior symptoms who are ACE inhibitor intolerant, unless contraindicated, to reduce morbidity and mortality.108,345,415,450 (Level of Evidence: A)**

Class IIa

1. **ARBs are reasonable to reduce morbidity and mortality as alternatives to ACE inhibitors as first-line therapy for patients with HFrEF, especially for patients already taking ARBs for other indications, unless contraindicated.451–456 (Level of Evidence: A)**

Class IIb

1. **Addition of an ARB may be considered in persistently symptomatic patients with HFrEF who are already being treated with an ACE inhibitor and a beta blocker in whom an aldosterone antagonist is not indicated or tolerated.420,457 (Level of Evidence: A)**

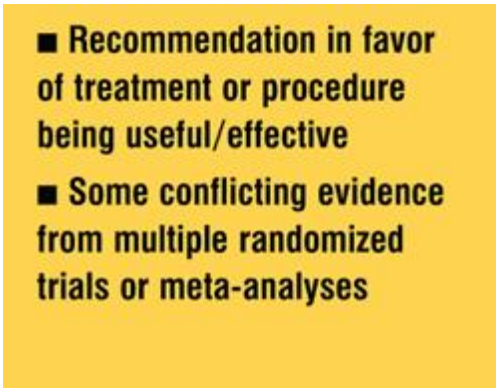
Class III: Harm

1. **Routine combined use of an ACE inhibitor, ARB, and aldosterone antagonist is potentially harmful for patients with HFrEF. (Level of Evidence: C)**

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

Class I, Level A =

Class II-A, Level A:



Class II-B, Level A:

Class III, Level C:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Table 1. Applying Classification of Recommendation and Level of Evidence

SIZE OF TREATMENT EFFECT

ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT	CLASS I *Benefit >>> Risk* Procedure/Treatment **SHOULD** be performed/ administered	CLASS IIa *Benefit >> Risk* *Additional studies with focused objectives needed* **IT IS REASONABLE** to perform procedure/administer treatment	CLASS IIb *Benefit ≥ Risk* *Additional studies with broad objectives needed; additional registry data would be helpful* Procedure/Treatment **MAY BE CONSIDERED**	CLASS III *No Benefit* or CLASS III *Harm* Procedure/ Test / Treatment COR III: No benefit — Not Helpful — No Proven Benefit COR III: Harm — Excess Cost w/o Benefit or Harmful — 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
Suggested phrases for writing recommendations	should is recommended is indicated is useful/effective/beneficial	is reasonable can be useful/effective/beneficial is probably recommended or indicated	may/might be considered may/might be reasonable usefulness/effectiveness is unknown/unclear/uncertain or not well established	COR III: No Benefit is not recommended is not indicated should not be performed/ administered/ other is not useful/ beneficial/ effective COR III: Harm potentially harmful causes harm associated with excess morbidity/mortality should not be performed/ administered/ other
Comparative effectiveness phrases†	treatment/strategy A is recommended/indicated in preference to treatment B treatment A should be chosen over treatment B	treatment/strategy A is probably recommended/indicated in preference to treatment B it is reasonable to choose treatment A over treatment B		

A recommendation with Level of Evidence B or C does not imply that the recommendation is weak. Many important clinical questions addressed in the guidelines do not lend themselves to clinical trials. Although randomized trials are unavailable, there may be a very clear clinical consensus that a particular test or therapy is useful or effective.

*Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use.

†For comparative effectiveness recommendations (Class I and IIa; Level of Evidence A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

http://circ.ahajournals.org/content/suppl/2013/06/04/CIR.0b013e31829e8776.DC1/Online_Data_Supplement.pdf

---

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Effectiveness of ACE/ARBs at reducing the risk of death and reducing hospitalization in patients with heart failure; which patients should receive ACE/ARB therapy; initiation and maintenance of therapy; and risks of treatment with ACE/ARB therapy.

1a.7.2. Grade assigned for the quality of the quoted evidence <u>with definition</u> of the grade:

Level A-

Level C-

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

See table in 1a.4.4

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 1993-2012

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

There were 2 meta-analyses, 10 randomized controlled trials, 3 comparative studies, and 1 review paper supporting the ACE/ARB recommendations.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

There are many solid randomized controlled trials that show that the benefits of using any ACE or ARB greatly outweigh the harms.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

ACE inhibitors can reduce the risk of death and reduce hospitalization in HF*r*EF. The benefits of ACE inhibition were seen in patients with mild, moderate, or severe symptoms of HF and in patients with or without CAD.

In several placebo-controlled studies, long-term therapy with ARBs produced hemodynamic, neurohormonal, and clinical effects consistent with those expected after interference with the renin-angiotensin system. Reduced hospitalization and mortality have been demonstrated. ACE inhibitors remain the first choice for inhibition of the renin-angiotensin system in systolic HF, but ARBs can now be considered a reasonable alternative.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

The majority of the adverse reactions of ACE inhibitors can be attributed to the 2 principal pharmacological actions of these drugs: those related to angiotensin suppression and those related to kinin potentiation. Other types of adverse effects may also occur (eg, rash and taste disturbances). Up to 20% of patients will experience an ACE inhibitor–induced cough. With the use of ACE inhibitors, particular care should be given to the patient's volume status, renal function, and concomitant medications (Sections 7.3.2.1 and 7.3.2.9). However, most HF patients (85% to 90%) can tolerate these drugs.

The risks of ARBs are attributed to suppression of angiotensin stimulation. These risks of hypotension, renal dysfunction, and hyperkalemia are greater when combined with another inhibitor of this neurohormonal axis, such as ACE inhibitors or aldosterone antagonists.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

The articles supporting the ACE and ARB recommendations were from 1993-2012. However, the overall literature search was through Oct, 2011, with select articles included through April, 2013.

We ran a search for Heart Failure and ACE ARB treatment for 2014 and 2015. There are only a few studies that are directly applicable to the target population; none would change the recommendation to use ACE/ARB therapy.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form NQF_measure_submission_evidence_HF_ACE_2.docx
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., the benefits or improvements in quality envisioned by use of this measure)* In the absence of contraindications, ACE inhibitors or ARBs are recommended for all patients with symptoms of heart failure and reduced left ventricular systolic function. ACE inhibitors remain the first choice for inhibition of the renin-angiotensin system in chronic heart failure, but ARBs can now be considered a reasonable alternative. Both pharmacologic agents have been shown to decrease the risk of death and hospitalization. Additional benefits of ACE inhibitors include the alleviation of symptoms and the improvement of clinical status and overall sense of well-being of patients with heart failure. 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 endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Data Source: 2013 is the most recent year for which PQRS Experience Report measure data is available. The average performance rates on Heart Failure (HF) – Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD) over the last several years are as follows:
2010: 85.6%
2011: 79.7%
2012: 82.2%
2013: 83.5%
2013 Small Group Practice Exception Rate: 1.3%
Reference: Center for Medicare and Medicaid Services. 2013 Reporting Experience Including Trends. Available: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/PQRS/
*It is important to note that PQRS was a voluntary reporting program, with approximately 51% of eligible professionals participating using any reporting option in 2013, and performance rates may not be nationally representative.

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.**
According to Fonarow and colleagues (2010), for aggregate practices at baseline, an ACEI/ARB was prescribed for 11 165 (79.8%) of 13 987 eligible patients.

http://circ.ahajournals.org/content/122/6/585.full
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 endorsement maintenance. 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 subcriterion on improvement (4b.1) 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 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
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

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR

- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Severity of illness
1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.
Heart Failure affects over 5.7 million Americans (2.4% in 2008; 2.7% in 2010) and that number is expected to rise consistently over the next 15 years (Heidenreich et al 2011, Mozaffarian 2015), The AHA forecasting study predicts that total costs for heart failure in the 18-44 age group will increase from $1.51 billion to $2.48 billion, while the costs for the 65-79 age group will increase from $11.50 billion to $29.9 billion (Heidenreich et al 2011).

In a 2014 article, Storrow et al writes that heart failure results in nearly 1 million annual hospital stays (Go 2013, Chen 2011), and is the top reason for Medicare hospital readmissions (Jencks 2009, Dharmarajan 2013). The vast majority of patients hospitalized for acute heart failure (AHF) are originally evaluated and managed in the emergency department (ED). Prior data suggest more than 80% of ED patients with AHF are admitted to the hospital and have a median inpatient length of stay (LOS) of approximately 3.4 days. Of the $39.2 billion dollars spent on heart failure care in the United States in 2010, hospital stay was the single largest proportion of this expenditure (AHA 2010, Heidenreich 2011). Among Medicare beneficiaries, hospital stay accounts for more than 50% of all heart failure costs in the last 6 months of life (Blecker 2012). Despite a small decline in the AHF hospital stay rate among Medicare beneficiaries over the last decade (Go 2013, Chen 2011), mortality remains high (Chen 2011, Richardson 2002, Roger 2004) and uneven across

Storrow et al (2014) utilized Nationwide Emergency Department Sample AHF data from 2006 to 2010 to describe admission proportion, hospital length of stay (LOS), and ED charges as a surrogate for resource utilization. Results were compared across U.S. regions, patient insurance status, and hospital characteristics. They concluded that a very high proportion of ED patients with AHF are admitted nationally, with significant variation in disposition and procedural decisions based on region of the country and type of insurance, even after adjusting for potential confounding.

1c.4. Citations for data demonstrating high priority provided in 1a.3
AHA Writing Group Members , Lloyd-Jones D., Adams R.J., Brown T.M., et al; Heart disease and stroke statisticsd2010 update: a report from the American Heart Association. Circulation. 2010;121:e46-215.
Blecker S., Herbert R., Brancati F.L.; Comorbid diabetes and end-of-life expenditures among Medicare beneficiaries with heart failure. J Card Fail. 2012;18:41-46.
Chen J., Normand S.L., Wang Y., Krumholz H.M.; National and regional trends in heart failure hospital stay and mortality rates for Medicare beneficiaries, 1998–2008. JAMA. 2011;306:1669-1678.
Dharmarajan K., Hsieh A.F., Lin Z., et al; Diagnoses and timing of 30-day readmissions after hospital stay for heart failure, acute myocardial infarction, or pneumonia. JAMA. 2013;309:355-363.
Go A.S., Mozaffarian D., Roger V.L., et al; Heart disease and stroke statistics—2013 update: a report from the American Heart Association. Circulation. 2013;127:e6-e245.
Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, Finkelstein EA, Hong Y, Johnston SC, Khera A, Lloyd-Jones DM, . Nelson SA, Nichol G, Orenstein D, Wilson PWF, Woo YJ on behalf of the American Heart Association Advocacy Coordinating Committee Stroke Council Council on Cardiovascular Radiology and Intervention Council on Clinical Cardiology Council on Epidemiology and Prevention Council on Arteriosclerosis Thrombosis and Vascular Biology Council on Cardiopulmonary Critical Care Perioperative and Resuscitation Council on Cardiovascular Nursing Council on the Kidney in Cardiovascular Disease Council on Cardiovascular Surgery and Anesthesia, and Interdisciplinary Council on Quality of Care and Outcomes Research. Forecasting the Future of Cardiovascular Disease in the United States: A Policy Statement From the American Heart Association. Circulation, March 1, 2011 vol. 123 no. 8 933-944. http://circ.ahajournals.org/content/123/8/933/T1.expansion.html

Jencks S.F., Williams M.V., Coleman E.A.; Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009;360:1418-1428.
Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation 2015;131:e29–322. doi:10.1161/CIR.0000000000000152
Richardson L.D., Asplin B.R., Lowe R.A.; Emergency department crowding as a health policy issue: past development, future directions. Ann Emerg Med. 2002;40:388-393.

Roger V.L., Weston S.A., Redfield M.M., et al; Trends in heart failure incidence and survival in a community-based population. JAMA. 2004;292:344-350.

Storrow AB, MD; Jenkins CA; Self WH; Alexander PT; Barrett TW; Han JH; McNaughton CD; Heavrin BS; Gheorghiade M; Collins SP. The Burden of Acute Heart Failure on U.S. Emergency Departments. JCHF. 2014;2(3):269-277. doi:10.1016/j.jchf.2014.01.006

1c.5. IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable. Not a PRO-PM.

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

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.)*
Centers for Medicare & Medicaid Services (CMS) eCQM Library webpage at: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/eCQM_Library.html Value set details at VSAC webpage: 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:** EP_CMS135v4_NQF0081_HF_ACEARB.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:** 0081_AMAPCPI_HF-ACEARB_ValueSets_June2015-635712727320959997.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Supporting guidelines and coding value sets included in the measure are reviewed on an annual basis. This annual review resulted in very limited changes to adhere to current eCQM industry standards and preserve the original measure intent.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who were prescribed* ACE inhibitor or ARB therapy either within a 12 month period when seen in the outpatient setting or

at hospital discharge

*Prescribed may include:
Outpatient setting: prescription given to the patient for ACE inhibitor or ARB therapy at one or more visits in the measurement period OR patient already taking ACE inhibitor or ARB therapy as documented in current medication list
Inpatient setting: prescription given to the patient for ACE inhibitor or ARB therapy at discharge OR ACE inhibitor or ARB therapy to be continued after discharge as documented in the discharge medication list

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
At least once during the 12 consecutive month measurement period when seen in the outpatient setting OR at each hospital discharge during the 12 consecutive month measurement period if seen in the inpatient setting.

S.6. **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, 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.
For EHR:
HQMF eMeasure developed and is included in this submission.

For Registry:
Definitions:
Prescribed – Outpatient setting: May include prescription given to the patient for ACE inhibitor or ARB therapy at one or more visits in the measurement period OR patient already taking ACE inhibitor or ARB therapy as documented in current medication list.
Prescribed – Inpatient setting: May include prescription given to the patient for ACE inhibitor or ARB therapy at discharge OR ACE inhibitor or ARB therapy to be continued after discharge as documented in the discharge medication list.

Report CPT Category II Code, 4010F : Angiotensin converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB) therapy prescribed or currently being taken

S.7. **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.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
For EHR:
HQMF eMeasure developed and is included in this submission.

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

DENOMINATOR NOTES:
To meet this measure, it must be reported for all heart failure patients a minimum of once during the measurement period when seen in the outpatient setting AND reported at each hospital discharge during the measurement period.

The requirement of “Count >=2 of Encounter, Performed“ is to establish that the eligible professional has an existing relationship with the patient.

For Registry:
Option 1, Outpatient Setting:
Patients aged >= 18 years
AND
Diagnosis for heart failure (ICD-9-CM) [for use 1/1/2015-9/30/2015]: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9
Diagnosis for heart failure (ICD-10-CM) [for use 10/01/2015-12/31/2015]: I11.0, I13.0, I13.2, I50.1, I50.20, I50.21, I50.22, I50.23, I50.30, I50.31, I50.32, I50.33, I50.40, I50.41, I50.42, I50.43, I50.9
AND
Patient encounter(s) during reporting period (CPT): 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99304, 99305, 99306, 99307, 99308, 99309, 99310, 99324, 99325, 99326, 99327, 99328, 99334, 99335, 99336, 99337, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350
AND
Two Denominator Eligible Visits
AND
Left ventricular ejection fraction (LVEF) < 40% or documentation of moderately or severely depressed left ventricular systolic function: 3021F

Option 2, Inpatient Setting:
Patients aged >= 18 years
AND
Diagnosis for heart failure (ICD-9-CM) [for use 1/1/2015-9/30/2015]: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9
Diagnosis for heart failure (ICD-10-CM) [for use 10/01/2015-12/31/2015]: I11.0, I13.0, I13.2, I50.1, I50.20, I50.21, I50.22, I50.23, I50.30, I50.31, I50.32, I50.33, I50.40, I50.41, I50.42, I50.43, I50.9
AND
Patient encounter during reporting period (CPT): 99238, 99239
AND
Left ventricular ejection fraction (LVEF) < 40% or documentation of moderately or severely depressed left ventricular systolic function: 3021F

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Documentation of medical reason(s) for not prescribing ACE inhibitor or ARB therapy (eg, hypotensive patients who are at immediate risk of cardiogenic shock, hospitalized patients who have experienced marked azotemia, allergy, intolerance, other medical reasons)
Documentation of patient reason(s) for not prescribing ACE inhibitor or ARB therapy (eg, patient declined, other patient reasons)
Documentation of system reason(s) for not prescribing ACE inhibitor or ARB therapy (eg, other system reasons)

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
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. This measure was developed using PCPI exception methodology which 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 : Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction, exceptions may include medical reasons (e.g. hypotensive patients who are at immediate risk of cardiogenic shock, hospitalized patients who have experienced marked azotemia), patient, and/or system reasons for not prescribing an ACE/ARB. Where examples of exceptions are included in the measure language, value sets for these examples are developed and included in the eSpecifications. 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.

Additional details by data source are as follows:

For EHR:
HQMF eMeasure developed and is included in this submission.

For Registry:
Append a modifier to CPT Category II Code:
4010F-1P : Documentation of medical reason(s) for not prescribing angiotensin converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB) therapy (eg, hypotensive patients who are at immediate risk of cardiogenic shock, hospitalized patients who have experienced marked azotemia, allergy, intolerance, other medical reasons)

4010F-2P : Documentation of patient reason(s) for not prescribing angiotensin converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB) therapy (eg, patient declined, other patient reasons)

4010F-3P : Documentation of system reason(s) for not prescribing angiotensin converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB) therapy (eg, other system reasons)

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 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.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
No risk adjustment or risk stratification

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)

S.16. Type of score:
Rate/proportion
If other:

S.17. 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.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*
To calculate performance rates:

1. Find the patients who meet the initial population (ie, 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 (ie, 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 (ie, 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: Documentation of medical reason(s) for not prescribing ACE inhibitor or ARB therapy (eg, hypotensive patients who are at immediate risk of cardiogenic shock, hospitalized patients who have experienced marked azotemia); Documentation of patient reason(s) for not prescribing ACE inhibitor or ARB therapy; Documentation of system reason(s) for not prescribing ACE inhibitor or ARB 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 (ie, 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.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable. The measure is not based on a sample.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. The measure is not based on a survey.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Patient eligibility is determined by a set of defined criteria relevant to a particular measure. If data required to determine patient eligibility are missing, those patients/cases would be ineligible for inclusion in the denominator and therefore the patient/case would be deleted.

If data required to determine if a denominator eligible patient qualifies for the numerator (or has a valid exclusion/exception) are missing, this case would represent a quality failure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
not applicable

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Clinician : Group/Practice, Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Home Health, Hospital/Acute Care Facility, Other, Post Acute/Long Term Care Facility : Long Term Acute Care Hospital, Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility
If other: Domiciliary

S.28. **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.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_0081_Heart_Failure_-HF-_-_Angiotensin-Converting_Enzyme_Inhibitor_or_ARB_Therapy_for_LVSD_Testing_Attachment.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: Heart Failure (HF) – Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)
Date of Submission: Click here to enter a date

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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.

2b2. 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.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

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.

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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/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*).

Data 1 (EHR - Validity Against the Gold Standard)

The data source is EHR data.

Bonnie Patient Test Deck

As a supplement to the EHR reliability testing performed on this measure, a deck of patient test cases have 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 2 (GPRO Registry)

The data source is the Centers for Medicare & Medicaid Services (CMS) PQRS GPRO database.

Data 3 (EHR – Exceptions Analysis)

The data source is EHR data.

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

Data 1 (EHR - Validity Against the Gold Standard)

The data are collected from patients sampled from 2007.

Data 2 (GPRO Registry)

The data are for the time period January 2013 – December 2013, and cover the entire United States.

Data 3 (EHR – Exceptions Analysis)

The data are collected from patients sampled from 2009.

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.26***)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

Data 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 Registry)

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 2 (GPRO Registry)

The total number of physicians reporting on this measure is 3,728. Of those, 1,244 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, 33.4 percent of physicians are included in the analysis, and the average number of quality reporting events is 31.5 for a total of 39,242 events. The range of quality reporting events for 1,244 physicians included is from 319 to 10. The average number of quality reporting events for the remaining 66.6 percent of physicians who aren't included is 3.2.

Data 3 (EHR – Exceptions Analysis)

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

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

Data 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 Registry)

There were 39,242 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 3 (EHR – Exceptions Analysis)

The sample consisted of approximately 127 eligible patients.

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

Data 1 (EHR - Validity Against the Gold Standard)

The data sample was used for the purposes of reliability and validity testing.

Data 2 (GPRO Registry)

The same data sample was used for reliability testing and exceptions analysis.

Face Validity (Data 2)

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 3 (EHR – Exceptions Analysis)

The data sample was used for the exception analysis only.

1.8. What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Data 1 (EHR - Validity Against the Gold Standard)

Patient-level socio-demographic (SDS) variables were not captured as part of the testing.

Data 2 (GPRO Registry)

Patient-level socio-demographic (SDS) variables were not captured as part of the testing.

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*)

Data 1 (EHR - Validity Against the Gold Standard)

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

Data 2 (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.

2a2.3. For each level 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*)

Data 1 (EHR - Validity Against the Gold Standard)

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

Data 2 (GPRO Registry)

For this measure, the reliability at the minimum level of quality reporting events (10) was 0.83. The average number of quality reporting events for physicians included is 31.5. The reliability at the average number of quality reporting events was 0.94.

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?)*

Data 1 (EHR - Validity Against the Gold Standard)

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

Data 2 (GPRO Registry)

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

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. For each level 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)*

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)

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

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Data 1 (EHR - Validity Against the Gold Standard)

Of the 254 patients sampled, automated EHR review detected 217 (85.4%) with an active electronic prescription for an ACE inhibitor or ARB. Of the remaining 37 patients, 23(62.2%) met one or more of the exclusion criteria. Performance on the ACE inhibitor and ARB quality measure was 93.9% by using automated EHR review.

Among the 14 patients without an active prescription for an ACE inhibitor or an ARB in the EHR, manual review of clinicians' notes in the EHR revealed that 5 patients had been prescribed an ACE inhibitor or ARB that was not recorded in the medication list. Six patients were found to have exclusion criteria through manual chart review. In addition, two patients met the exclusion criteria on automated review, but upon manual review, the exceptions were found to be false.

Performance on the measure was calculated to be 98.7% through comparison of automated and manual EHR review.

Face Validity (Data 2)

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

2b2.4. 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?*)

Data 1 (EHR - Validity Against the Gold Standard)

The automated quality assessment had a sensitivity of 97.7% for identifying patients with heart failure taking an ACE inhibitor or ARB. The automated quality assessment captured 21 of 29 patients with valid exclusion criteria (sensitivity, 72.4%), and 2 of 23 patients who met exclusion criteria were judged not to have a true exclusion.

Data 2 (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% 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)

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

Data 1 (EHR - Validity Against the Gold Standard)

This data sample was not used to test exclusions.

Data 2 (GPRO Registry)

With the information available from the GPRO Registry, 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 3 (EHR-Exceptions Analysis)

Exceptions included documentation of medical reason(s), patient reason(s) and system reason(s) for not prescribing ACE inhibitor or ARB therapy. Exceptions were analyzed for frequency and variability across providers.

2b3.2. 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)

Data 1 (EHR - Validity Against the Gold Standard)

This data sample was not used to test exclusions.

Data 2 (GPRO Registry)

Amongst the 1244 physicians with the minimum (10) number of quality reporting events, there were a total of 8,056 exceptions reported. The average number of exceptions per physician in this sample is 6.5. The overall exception rate is 17.03%.

Data 3 (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 127 exceptions revealed 99.5% of exceptions were medical reasons for not prescribing ACE inhibitor or ARB therapy. Medical reason exceptions consisted of clinical contraindications, drug allergy and drug intolerance.

2b3.3. 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*)

Exceptions are necessary to account for those situations when it is not medically appropriate to prescribe ACE inhibitor or ARB 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 ACE inhibitor or ARB 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.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

Not applicable

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. 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)*

Not applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis*

was used)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

Not applicable

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. 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*)

Not applicable

***2b4.11. 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*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

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 Registry)

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

2b5.2. 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*)

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 Registry)

Based on the sample of 1,244 included physicians, the mean performance rate is 0.80, the median performance rate is 0.94 and the mode is 1.00. The standard deviation is 0.29. The range of the performance rate is 1.00, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.29 (0.71 - 1.00).

2b5.3. 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?*)

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 Registry)

The range of performance from 0.00 to 1.00 suggests there's clinically meaningful variation across physicians' performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

This test was not performed for this measure.

2b6.2. 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*)

This test was not performed for this measure.

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

This test was not performed for this measure.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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)

Data are not available to complete this testing.

2b7.2. 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)

Data are not available to complete this testing.

2b7.3. 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)

Data are not available to complete this testing.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Generated 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)*

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.**

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.**

Attachment **Attachment:** NQF_0081_Feasibility_Scorecard_Bonnie_Output_Screen_Shots_Revised.pdf

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

We have not identified any 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.

Planned	Current Use (for current use provide URL)
	Public Reporting Physician Quality Rating System http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/pqrs/index.html Payment Program Meaningful Use Stage II http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Stage_2.html Quality Improvement with Benchmarking (external benchmarking to multiple organizations) PINNACLE Registry http://cvquality.acc.org/en/NCDR-Home/Registries/Outpatient-Registries.aspx

4a.1. **For each CURRENT use, checked above, provide:**

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

1) Physician Quality Reporting System (PQRS)-Sponsored by the Centers for Medicare and Medicaid Services (CMS)

Purpose: PQRS is a national reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs). The program provides an incentive payment to practices with EPs (identified on claims by their individual National Provider Identifier [NPI] and Tax Identification Number [TIN]). Eps satisfactorily report data on quality measures for covered Physician Fee Schedule (PFS) services furnished to Medicare Part B Fee-for-Service (FFS) beneficiaries (including Railroad Retirement Board and Medicare Secondary Payer). Beginning in 2015, the program also applies a payment adjustment to EPs who do not satisfactorily report data on quality measures for covered professional services. Source: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html It is our understanding that CMS is also planning to move towards publicly reporting physician data via Physician Compare.

2) Meaningful Use Stage 2 (EHR Incentive Program) – Sponsored by the Centers for Medicare and Medicaid Services (CMS)

The Medicare and Medicaid EHR Incentive Programs provide incentive payments to eligible professionals, eligible hospitals, and

critical access hospitals (CAHs) as they adopt, implement, upgrade or demonstrate meaningful use of certified EHR technology.

These professionals are eligible for incentive payments for the "meaningful use" of certified EHR technology, if all program requirements are met, including successful implementation and reporting of program measures, which include this measure, to demonstrate meaningful use of EHR technology.

3) PINNACLE Registry (URL: http://cvquality.acc.org/en/NCDR-Home/Registries/Outpatient-Registries.aspx)
The PINNACLE Registry® is cardiology´s largest outpatient quality improvement registry, capturing data on coronary artery disease, hypertension, heart failure and atrial fibrillation. The PINNACLE Registry® continues to grow rapidly, with more than 2400 providers representing almost 800 unique office locations across the U.S submitting data to the registry as of the fourth quarter of 2013. As of the fourth quarter of 2013, the registry has more than 13 million patient encounter records. PINNACLE assists practices in understanding and improving care through the production and distribution of quarterly performance reports. These reports, covering all valid patient encounters, detail adherence to 28 cardiovascular clinical measures at the physician, location, and practice levels across coronary artery disease, hypertension, heart failure and atrial fibrillation. All jointly developed ACC/AHA/PCPI performance measures for these topics are reported by the registry.

4a.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 new insurance marketplace.

4a.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. 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.

4b. 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.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

2013 PQRS Experience Report*:
Data Source: 2013 is the most recent year for which PQRS Experience Report measure data is available. The average performance rates on Heart Failure (HF) – Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD) over the last several years are as follows:

2010: 85.6%
2011: 79.7%
2012: 82.2%
2013: 83.5%

2013 Small Group Practice Exception Rate: 1.3%
Reference: Center for Medicare and Medicaid Services. 2013 Reporting Experience Including Trends. Available: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/PQRS/

*It is important to note that PQRS was a voluntary reporting program, with approximately 51% of eligible professionals participating using any reporting option in 2013, and performance rates may not be nationally representative.

4b.2. **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.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We are not aware of any unintended consequences at this time, but we take unintended consequences very seriously and therefore continuously monitor to identify actions that can be taken to mitigate them.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

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

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

5a. Harmonization

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 completely harmonized?

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

5b. Competing Measures

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

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

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** AMA-PCPI
Co.2 **Point of Contact:** Caryn, Davidson, caryn.davidson@ama-assn.org, 312-464-4465-
Co.3 **Measure Developer if different from Measure Steward:** AMA-PCPI
Co.4 **Point of Contact:** Caryn, Davidson, pcpimeasures@ama-assn.org, 312-464-4465-

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 and ACC/AHA measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study are invited to participate as equal contributors to the measure development process. In addition, the PCPI and ACC/AHA strive to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

Work Group members:
Craig T. Beam, CRE (patient representative)
Ileana L. Piña, MD, FACC (cardiology, heart failure)
Kathleen Blake, MD (cardiac electrophysiology)
Paul D. Rockswold, MD, MPH (family medicine)
Donald E. Casey, Jr., MD, MPH, MBA, FACP, FAHA (internal medicine)
Lawrence B. Sadwin (patient representative)
Sarah J. Goodlin, MD (geriatrics, palliative medicine)
Joanna D. Sikkema, MSN, ANP-BC, FAHA (cardiology)
Kathleen L. Grady, PhD, APN, FAAN, FAHA (cardiac surgery)
Carrie A. Sincak, PharmD, BCPS (pharmacy)
Randal F. Hundley, MD, FACC (cardiology, health plan representative)
John Spertus, MD, MPH (cardiology)
Mariell Jessup, MD, FACC, FAHA, FESC (cardiology, heart failure)
Patrick J. Torcson, MD, FACP, MMM (hospital medicine)
Thomas E. Lynn, MD (family medicine, measure implementation)
Elizabeth Torres, MD (internal medicine)
Frederick A. Masoudi, MD, MSPH (cardiology)

Mark V. Williams, MD, FHM (hospital medicine)
David Nilasena MD, MSPH, MS (general preventive medicine, public health, measure implementation)
John B Wong, MD (internal medicine)

American College of Cardiology Foundation
Charlene L. May
Melanie Shahriary, RN, BSN

American Heart Association
Cheryl Perkins, MD, RPh
Mark D. Stewart, MPH
Gayle Whitman, PhD, RN, FAHA, FAAN

American College of Cardiology Foundation/American Heart Association
Jensen S. Chiu, MHA
National Committee for Quality Assurance (NCQA) Liaison
Manasi Tirodkar, PhD, MS

The Joint Commission Liaison
Millie J. Perich, MS, RN

American Medical Association
Mark Antman, DDS, MBA
Heidi Bossley, MSN, MBA
Christopher Carlucci, MBA
Kerri Fei, MSN, RN
JoeAnn Jackson, MJ
Kendra Hanley, MS
Karen Kmetik, PhD
Pamela O'Neil, MPH
Samantha Tierney, MPH
Temaka Williams, MPH, MBA
Greg Wozniak, 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:** 12, 2014

Ad.4 **What is your frequency for review/update of this measure?** Coding/Specifications updates occur annually.

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2015

Ad.7 **Disclaimers:** Physician Performance Measures (Measures) and related data specifications have been developed by the American Medical Association (AMA) - convened Physician Consortium for Performance Improvement(R) (PCPI[R]), American College of Cardiology (ACC) and American Heart Association (AHA). These Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Measures require a license agreement between the user and the AMA, (on behalf of the PCPI), ACC or AHA. Neither the AMA, ACC, AHA, PCPI nor its members shall be responsible for any use of the Measures.

THE MEASURES 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 disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT[R]) or other coding contained in the specifications.

Ad.8 **Additional Information/Comments:**

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 #: 0229
Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older
Measure Steward: Centers for Medicare & Medicaid Services (CMS)
Brief Description of Measure: The measure estimates a hospital-level 30-day risk-standardized mortality rate (RSMR). Mortality is defined as death for any cause within 30 days after the date of admission for the index admission, for patients 18 and older discharged from the hospital with a principal diagnosis of heart failure (HF). CMS annually reports the measure for patients who are 65 years or older and are either Medicare fee-for-service (FFS) beneficiaries and hospitalized in non-federal hospitals or patients hospitalized in Veterans Health Administration (VA) facilities.
Developer Rationale: The goal of this measure is to improve patient outcomes by providing patients, physicians, hospitals, and policy makers with information about hospital-level, risk-standardized mortality rates following hospitalization for HF. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This measure was developed to identify institutions' whose performance is better or worse than would be expected based on their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

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

Numerator Statement: The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days of the index admission date for patients 18 and older discharged from the hospital with a principal diagnosis of HF.
Denominator Statement: This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 18 years or older. We have explicitly tested the measure in both age groups. The cohort includes admissions for patients aged 18 years and older discharged from the hospital with a principal discharge diagnosis of HF and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients 65 years and older who are either Medicare FFS beneficiaries admitted to non-federal hospitals or patients admitted to VA hospitals. Additional details are provided in S.9 Denominator Details.
Denominator Exclusions: The mortality measures exclude index admissions for patients:
1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility.
2. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission;
4. Discharged against medical advice (AMA); or
5. Patients undergoing LVAD implantation or heart transplantation during an index admission or who have a history of LVAD or heart transplant in the preceding year.

For patients with more than one admission for a given condition in a given year, only one index admission for that condition is

randomly selected for inclusion in the cohort.

For Medicare FFS patients, the measure additionally excludes admissions for patients without at least 30 days post-discharge enrollment in FFS Medicare (because the 30-day mortality outcome cannot be assessed in this group).

Measure Type: Outcome
Data Source: Administrative claims, Other, Paper Medical Records
Level of Analysis: Facility
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

IF this measure is paired/grouped, NQF#/title: 0330: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

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

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: 5/9/07 **Most Recent Endorsement Date:** 1/18/12

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

Public and Member Comments on original measure

- Given the advanced age of many HF patients, many in palliative care programs, many deaths cannot be considered a result of substandard care.

Committee response:

- Patients in hospice care are excluded and risk factors account for frailty.

Public and Member Comments on revised measure:

- Clarify the data sources that were used in the all payer data testing.

Developer Response:

- The data source used to complete the all payer testing was the state of California's Patient Discharge Database (PDD) which contains records for all discharges from all non-Federal hospitals located in California. California is a diverse state, and, with more than 37 million residents, California represents 12% of the US population. In 2006, there were approximately 3 million adult discharges from more than 450 hospitals. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations and to evaluate rates of both readmission and mortality. Specifically, patients from this database are linked to the California Death Statistical Master File (DSMF) using social security number in order to validate and record deaths.

Steering Committee: Agreed that the developer answered the comment.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing a rationale that supports the relationship of the outcome to at least one healthcare structure, process, intervention, or service.

- This measure calculates hospitals' 30-day risk-standardized mortality rate for patients who have been

hospitalized with heart failure (HF).

- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of complications, provision of evidenced-based care, discharge planning, management of care transitions, medication reconciliation, and patient education.
- The developer states that numerous studies show that appropriate and timely treatment for HF patients can reduce the risk of mortality within 30 days of hospital admission; that trials of interventions which improve patient education upon discharge have been shown to improve survival for HF patients; and that evidence that hospitals have been able to reduce mortality rates through these quality-of-care initiatives illustrates the degree to which hospital practices can affect mortality rates.
- The developer states that this measure was developed to identify institutions whose performance is better or worse than would be expected based on their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Question for the Committee:

- *Does the Committee agree that hospitals have the ability to influence 30-day mortality rates among heart failure patients?*

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

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

- The developer notes that studies suggest quality gaps in hospital care for heart failure patients, particularly in the transition to outpatient care.
- The developer also notes that there is substantial inter-hospital variation in the risk of death among heart failure patients that is not clearly explained by differences in case mix, and suggests that measurement of this patient outcome allows for a broad view of quality of care.
- The developer provides performance data from four measurement periods, covering a total of 991,007 admissions.
- The data show that the average 30-day risk-standardized HF mortality rate was **11.7 percent** during the measurement period of 07/2011–06/2014.
- During this same period, scores ranged from a **minimum of 7.0%** to a **maximum of 19.3%**, with the **10th percentile at 10.1%** and the **90th percentile at 13.4%.**
- To help in assessment of potential disparities, the developers also provide performance scores for hospitals serving a low proportion of Medicaid patients vs. those serving a high proportion of Medicaid patients, and performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients.
- By proportion of **Medicaid** patients:

  // **Low proportion (=7.0%) Medicaid patients** // **High proportion (=29.6%) Medicaid patients**
  Number of Measured Entities (Hospitals)// 373 // 373
  Number of Patients// 84,068 patients in low-proportion hospitals/74,416 in high-proportion hospitals
  Maximum// 15.1 // 16.4
  90th percentile// 13.2 // 13.3
  75th percentile// 12.4 // 12.4
  Median (50th percentile)// 11.6 // 11.3
  25th percentile// 10.9 // 10.5
  10th percentile// 10.1 // 9.5
  Minimum // 7.7 // 7.2

- By proportion of **African-American** patients:

  // **Low Proportion (=0%) Af-Am patients** // **High proportion (=23.3%) Af-Am patients**
  Number of Measures Entities (Hospitals)// 526 // 376

Number of Patients//31,904 patients in low-proportion hospitals/92,579 in high-proportion hospitals
Maximum// 16.4 // 15.9
90th percentile// 13.8 // 12.9
75th percentile// 12.8 // 12.0
Median (50%)// 12.0 // 11.1
25th percentile// 11.3 // 10.2
10th percentile// 10.8 // 9.3
Minimum// 9.1 // 7.2

- The developers do not provide interpretation or analysis of these data; in both cases there are differences in performance between the groups compared, but these differences do not appear to be substantial.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - Evidence is strong.
   - The measure steward provided a sound rationale to support this risk adjusted outcome measure, one that estimates a hospital-level 30 day risk standardized mortality after discharge from the hospital with a principal diagnosis of HF. More specifically the measure developed provided a very nice diagram showcasing the linkage between the health outcome, in this case decreased risk of mortality and the processes, interventions, or services that influence. Furthermore, they provide numerous studies that demonstrate this link based upon well designed studies.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- Strong.

1b. ***Committee's Comments on Performance Gap:***

- Data show that the average 30-day risk-standardized HF mortality rate was 11.7 percent during the measurement period of 07/2011–06/2014. During this same period, scores ranged from a minimum of 7.0% to a maximum of 19.3%, with the 10th percentile at 10.1% and the 90th percentile at 13.4%. So there is a clear gap.

1c. ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- This measure calculates 30-day all-cause mortality for patients hospitalized with heart failure (HF) using a risk standardized mortality ratio (RSMR), which is calculated as the ratio of the number of "predicted" to the number of "expected" deaths, multiplied by the national unadjusted mortality rate.

- The denominator includes patients aged 18 years and older discharged from the hospital with a principal discharge diagnosis of HF and with a complete claims history for the 12 months prior to admission. The measure can also be calculated for patients aged 65 and older only.
- The numerator includes patients who died of any cause within 30 days of the date of admission of the index HF hospitalization.
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.
- The numerator population is defined using vital status data, which may be derived from the Medicare Enrollment Database (EDB), State-based data systems, the Social Security Administration's Death Master File (DMF), or the Centers for Disease Control and Prevention's National Death Index (NDI).
- The data sources for this measure may include Medicare Part A and B claims, Veterans Health Administration claims, the Medicare Enrollment Database (EDB), and all-payer data sources such as the California Patient Discharge Database.
- The measure's time window can be specified from one to three years.
- The measure is risk-adjusted using a statistical risk model (see details below).

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- The developer has assessed reliability at both the data element and the performance score levels.
- **Data element reliability**:
  - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS. [*Note: NQF does not typically consider temporal consistency to be a valid method of demonstrating reliability of data elements*.]
  - In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.
  - Summarizing the results of this analysis, the developer notes that the frequency of some model variables increased between 2011 and 2014, which may reflect increased co-morbidity rates, but may also be due to increased coding opportunities on administrative claims.
  - The developer states that examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across three years.
- **Performance score reliability**:
  - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
  - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. [*Note: NQF considers this to be an appropriate method of assessing reliability*.]
  - A total of 991,007 admissions over a 3-year period were examined, with 494,297 in one sample and 496,710 in the other randomly-selected sample; two risk-standardized mortality rates (RSMR) were calculated for each hospital, one from each of the two separate samples.
  - The agreement between the two RSMRs for each hospital (as measured by intra-class correlation coefficient (ICC)) was **0.55**; the developers state that according to the conventional interpretation, this is considered a "moderate" level of agreement.
  - The developer notes that this analysis was limited to hospitals with 12 or more cases in each split sample, and

that splitting the total population into two samples resulted in a sample equivalent of only 1.5 years of data, whereas the measure is reported with the full three years of data. [*Note: It is unclear whether the measure itself is limited to hospitals with 12 or more cases; if it is not, then testing was not conducted with the measure as specified*.]
    - The developer expects that the correlation coefficient would be higher using a full three-year sample since it would include more patients.
- The developer's overall interpretation of reliability testing results is that the stability of the risk factor frequencies and odds ratios over time suggests that the underlying data elements are reliable, and that the ICC score from performance score analysis demonstrates moderate agreement across samples using a conservative approach to assessment.

Questions for the Committee:

- *Do the testing results presented by the developer demonstrate an adequate level of reliability?*
- *In addition to the consistency of measurement results, assessments of performance score reliability often examine the ability of the measure to differentiate between measured entities. Do the reliability testing results reported by the developer demonstrate that meaningful differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- This measure calculates 30-day all-cause mortality for patients hospitalized with heart failure (HF).
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of complications, provision of evidenced-based care, discharge planning, management of care transitions, medication reconciliation, and patient education.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- The developer conducted empirical validity testing of the measure score.
- To assess validity, the developer compared scores from the administrative claims-based measure (i.e., the measure as specified) to scores derived from medical record review in the same patient cohort.
- This assessment was conducted on data from 1998-2001, comprising 46,700 heart failure hospitalizations; the unadjusted 30-day mortality rate in this population was 11.9%.
- Hospital-level risk-standardized mortality rates (RSMRs) were estimated using the claims-based model and the medical record-based model; the linear relationship between these two sets of estimates was then examined, using regression techniques and weighting by the total number of cases in each hospital.
- The correlation between the claims-based RSMRs and the record-based RSMRs was estimated at 0.95, which the developer suggests shows that the resulting measure from the administrative claims model is as good as that from the medical record model.

Questions for the Committee:

- *Do the method and results of testing demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients in the following categories are excluded from the measure:
  1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility.
  2. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
  3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission;
  4. Discharged against medical advice (AMA); or
  5. Patients undergoing LVAD implantation or heart transplantation during an index admission or who have a history of LVAD or heart transplant in the preceding year.
- Additional exclusions listed by the developer include:
  - HF admissions within 30 days of discharge from a qualifying index admission, which are identified by comparing the discharge date from the index admission with the readmission date
  - For Medicare FFS patients, the measure additionally excludes admissions for patients without at least 30 days post-discharge enrollment in FFS Medicare (because the 30-day mortality outcome cannot be assessed in this group)
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- The number and percentage of patients excluded for each criterion are as follows:
  1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility: **88,023 (6.45%)**
  2. Inconsistent or unknown vital status or other unreliable demographic (age and gender) data: **55 (<0.01%)**
  3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission: **18,753 (1.37%)**
  4. LVAD or Transplant in index admission or prior year: **2,362 (0.17%)**
  5. Discharged against medical advice (AMA): **5,933 (0.43%)**
- The developer also provides the distribution across hospitals for each exclusion criterion.
- The developer notes that the first exclusion criterion accounts for the majority of all exclusions and is meant to ensure a clinically coherent cohort, preventing the inclusion of patients who likely did not have clinically significant HF.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- To control for differences in case mix between hospitals, the developers have risk-adjusted this measure using a statistical risk model.
- Specifically, the developer employs a hierarchical logistic regression model to create a hospital-level 30-day risk-standardized mortality ratio (RSMR).
- The developer notes that the risk-adjustment approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes both within and between hospitals.
- At the patient level, the odds of mortality are adjusted for age, sex, and selected clinical covariates present at the time of admission.
- For each patient, covariates were obtained from claims records extending 12 months prior to and including the index admission. The covariates are defined using condition categories (CCs), which are clinically-meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes.
- The measure does not adjust for CCs that were possible adverse events of care and that were only recorded in the index admission.
- Variables considered for inclusion in the risk-adjustment model were patient-level risk-adjustors that are expected to be predictive of mortality based on empirical analysis, prior literature, and clinical judgment.
- The final set of 24 clinical risk-adjustment variables, along with odds ratios for each, is available in the testing

attachment.

- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature, conceptual pathways, and feasibility of collection from national data sources.
- **Conceptual analysis of the need for SDS adjustment**:
    - The developers note that there is a large body of literature linking various SDS factors to worse health status and higher mortality over a lifetime, with income, education, and occupational level being the most commonly examined variables, though literature directly related to 30-day mortality after heart failure hospitalization is much more limited.
    - One potential pathway for SDS factors to affect 30-day mortality (independent of the quality of care) is patients' health status at the time of admission.
        - SDS factors can influence admission health status both due to the impact of multiple related stressors over a lifetime contributing to overall worse health, as well as through poorer access to care and potentially delayed presentation.
        - The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
    - Another potential pathway for SDS factors to affect 30-day mortality is for lower-income patients to elect not to follow prescribed care (for example, refill a prescription or keep a follow-up visit with a primary care provider) because limited resources create competing priorities for the patient.
    - The developers also argue that there are a number of pathways for SDS to affect 30-day mortality that are *not* independent of the quality of care. These may include:
        - Contextual effects, such as patients of low SDS presenting at lower-quality institutions for care;
        - Patients of low SDS receiving differentiated care as compared to counterparts of higher SDS—which may be appropriate or inappropriate in different instances.
- **Empirical analysis of SDS factors**:
    - The developers state that they found race (black vs non-black) and dual-eligible status—e.g. enrolled in both Medicare and Medicaid (obtained from CMS claims enrollment data)—to be the only two patient-level SDS variables available for direct examination. The conceptual relationship between race and mortality from heart failure was not explained by the developer, although they state that they "felt it was important to understand the association with race as well as more traditional socio-economic variables". Guidance from the NQF panel that examined use of SDS factors in risk-adjustment approaches noted that race should generally not be used as a proxy for socioeconomic status.
    - Also considered were a number of neighborhood-level variables (represented in a validated AHRQ composite index of SDS variables found in census data, including income and education) that could serve as a proxy for patient-level SDS.
        - These variables are linked to patients by zip code; however, the data are only linked at a **5-digit zip code level**—nine-digit zip code data, which may provide a more granular view of patient sociodemographic status, were not available.
        - Patients were identified as low SDS if they lived in a neighborhood in the lowest quartile of the AHRQ SDS index.
    - The developer's method was to first evaluate variation in the prevalence of low-SDS patients among providers; they then assessed the relationship (univariate) between the SDS variables and 30-day HF mortality, and examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
    - The developers' analysis found that the prevalence of SDS factors in the HF cohort does vary across measured entities.
        - With regard to the empirical association of each SDS variable with the outcome (univariate), the analysis found that patient-level observed HF mortality rates were lower for dual-eligible patients, black patients, and patients in the lowest AHRQ SDS quartile as compared to, respectively, mortality rates for non-dual-eligible patients, non-black patients, and patients above the lowest QHRQ SDS quartile.

    - With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that the effect size of each of these variables was small, that the c-statistic (i.e., predictive value) of the model was essentially unchanged with the addition of any of these variables, and that addition of the variables to the model had little to no effect on hospital performance.
- The developers note that, among the SDS variables that could be feasibly incorporated into their model, the relationship between minority status, dual-eligible status, and low SES (AHRQ indicator status) is in the opposite direction than what has been the expressed concern of stakeholders interested in adding such adjustment to the models (although it is unclear what, exactly, this means). The developers state that, given the controversial nature of incorporating such variables into a risk-model, they do not support doing so in a case that is unlikely to affect hospital profiling. As a result, the developers **did not incorporate SDS variables into this measure**.

- **Risk Model Diagnostics**
  - To assess the overall performance of their risk-adjustment model, the developers computed several summary statistics, including:
    - *Area under the receiver operating characteristic (ROC) curve* (also known as a ***c-statistic***, which reflects how well the risk model can distinguish those who have the outcome of interest from those who do not)
    - *Predictive ability* (the model's ability to distinguish high-risk subjects from low-risk subjects)
    - *Over-fitting indices (model calibration)* (to ensure that the model is not only describing the relationship between predictive variables and outcome in the development dataset but also providing valid predictions in new patients)
  - For the current measure cohort, the findings from this analysis are as follows:
    - C-statistic: **0.68**
      - A c-statistic of 0.68 means that for 68% of all possible pairs of patients—one who died and one who lived—the model correctly assigned a higher probability to those who died. Generally, a c-statistic of at least 0.70 is considered acceptable.
      - The developers interpret this as 'fair' model discrimination.
    - Predictive ability (lowest decile %, highest decile %): (**3.7, 27.3**)
      - The developers state that this indicates a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.
    - Overfitting indices (model calibration) [presented as (γ0, γ1)]:
      - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
        - 1st half of split sample: Calibration: (**0.0000, 1.0000**)
        - 2nd half of split sample: Calibration: (**-0.0035, 0.9928**)
  - The developer's overall interpretation of the results of their analyses is that the findings demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

- The developer also conducted additional analyses to determine whether the measure could be applied to Medicare FFS 65+ patients using only Medicare Part A data and whether it could be applied to a population of patients aged 18+ using all-payer data.
- The developers report that their results indicate their model had good discrimination and predictive ability in both groups.

Questions for the Committee:

- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk-adjustment variables present at the start of care?*

- *Does the Standing Committee agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*
- *Do you agree with the developer's decision, based on their empirical analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference:

- For public reporting of this measure, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate.
- If the RSMR's interval estimate does not include the national observed mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare website as "better than the U.S. national rate" or "worse than the U.S. national rate."
- If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or "the difference is uncertain."
- The developer reports that for the performance period of July 2011-June 2014, the median hospital RSMR was 11.7%, with a range of 7.0% to 19.3%. The interquartile range was 10.9%-12.4%.
- Out of 4,771 hospitals in the U.S., 145 performed "better than the U.S. national rate," 3,662 performed "no different from the U.S. national rate," and 93 performed "worse than the U.S. national rate." 871 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.
- The developer's interpretation of this data is that the variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for heart failure that support measurement to reduce the variation.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- While the developer did not decide to include SDS variables in their final model, they did compare measure results with and without SDS adjustment.

2b7. Missing Data

- N/A

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- Why are patients discharged alive on the day of admission or the following day excluded?
- In the Medicare population how can the vital status be unknown?
- Should not patients who are admitted with end stage heart failure and are sent out to Hospice be excluded?
- This 30-day all cause mortality measure for patients hospitalized with HF using a standardized risk standardized mortality ratio (RSMR) is well specified and has been updated to reflect the ICD9 CM CC map to capture all relevant co morbidities coded in patient administrative claims data. The measure developer also described other updates for the 2016 Reporting that notes the future exclusion of patients undergoing LVAD implantation or heart transplant during the index admission or who have had a history of LVAD or heart transplant in the previous year. The measure developer also provided the rationale behind that decision citing that this patient population would have distinctly different mortality risk and risk factors from the primary heart failure cohort.
- Denominator statement includes patients 18 and older discharged from hospital with a principal discharge diagnosis of HF and with a complete claims history for the 12 months prior to admission. The measure can also be calculated for patients 65 and older only as well. The denominator population was defined using ICD 9 and 10 codes, with all codes included in the submission.
- The numerator statement is "any patient who died of any cause within 30 days of the date of admission of the index HF hospitalization." Data used to construct numerator includes Death Master File, state-based data systems, and Medicare Enrollment Database, among others.
- Data sources used to construct this measure include the following: Medicare Part A and B claims, all payer data sources, Veterans Health Administration claims
- The measure can be calculated from 1 to three years.
- Based upon the rigor in defining all elements needed to construct the measures, and the incorporation of all important and relevant codes, I think that it is likely that this measures can be consistently implemented."

2a2.: ***Committee's Comments on Reliability-Testing:***

- I understand what they did and I am fine with it. But is 0.55 adequate?? And I am not sure I agree with the following: The developer expects that the correlation coefficient would be higher using a full three-year sample since it would include more patients. Without data to support this, I do not think such claims can be made.
- In keeping with the requirements for measure endorsement, the measure development assessed reliability at both the individual data element and at the performance score levels. Based upon the data submitted , the testing results demonstrate an adequate level of reliability. More specifically, the presentation of the odds ratios for each risk variable in the model remained relatively constant across the three years, even though some model variables increased during that same time frame. The measure developed cited that there may be an increase in co-morbidity rates or improved coding on administrative claims.
- The measure developer used a test-re-test approach to assess reliability of the hospital performance, a method endorsed by NQF. A total of 991,007 admissions over a 3 year period were examined, and RSMRs were calculated in one sample and in another randomly selected sample. The agreement between the two was .55 which was considered moderate level of agreement. Not nearly as high as I would have expected. Would love to see a re-run of the analysis with a more robust set of data over a longer time frame...

2b1.: ***Committee's Comments on Validity-Specifications***:

- Why are patients discharged alive on the day of admission or the following day excluded?
- In the Medicare population how can the vital status be unknown?
- Should not patients who are admitted with end stage heart failure and are sent out to Hospice be excluded?
- The numerator, denominator were appropriately defined. Specific data elements like 30-day all cause mortality.
- I think that the specifications are consistent with the evidence.

2b2.: ***Committee's Comments on Validity-Testing:***

- Excellent correlation between the claims-based RSMRs and the record-based RSMRs.
- The measure developed conducted empirical validity testing of the sure score. To assess validity, the measure developer compared scores from the administrative claims based measure to those computed from the medical record review from the same cohort of patients. The correlation between the claims based and the medical record review RSMR was estimated at 0.95 which suggests that the measure is as good as that which is derived from the medical record model.
- As such the methods and results of testing truly demonstrate sufficient validity so that accurate conclusions about quality can be made. Moreover, the score that is calculated form this measure is an indicator of quality.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Model is not great (c statistic is only 0.68).
- One of the variables they include in the model is Hf, but all pts have HF, so it does not make sense to include such a variable.
- The measure developer listed a number of exclusions and determined its impact by examining overall frequencies and proportions of the total cohort excluded for each exclusion criteria. The measure developer then presented the distribution of these exclusions across hospitals. Interesting to note that the majority of the excluded patient population were those who were discharged on the day of admission or the following day who were not transferred to another acute care facility, thus demonstrating that the patients included in the cohort are clinically relevant. Based upon the measure developers submission, a final set of clinical risk-adjustment variables, along with ORs were presented. The measure developers also considered a set of sociodemographic status (SDS) variables, presented a conceptual analysis for the need for SDS adjustment, and then set out to conduct an empirical analysis of the SDS factors. Based upon their findings, the measure developers first stated that the literature was limited in linking various SDS factors to 30 day mortality after HF. Howeer they did note that SDS could be feasibly incorporated into their model, however, noted that given the controversial nature of incorporating such variables into a risk model they've chosen not to do so.
- As for the risk adjustment model, the measure developers computed several summary statistics- C-statistic, predictive ability, and overfitting indices. While with a Cstatistic at .68, the measure developers noted that this is considered fair model discrimination, the predictive ability of distinguishing high risk subjects from low-risk subjects was good. Overall, the measure developer noted that their findings demonstrated that the risk adjustment model adequately controls for difference in case mix. Not being a statistician, I must defer to my colleagues who may have a different interpretation of the results. Stratifying for 65+ using Medicare part A data and for patients 18+ using all payer data, the model still had good discrimination.
- I think that the final variables in the risk model are adequately described. I believe that all of the risk adjustment variables are present at the start of care.

- Based upon the data presented around meaningful differences, there seems to be variability in the quality of care across hospitals. The median hospital RSMR was 11.7% with a range of 7% to 19.3%. The interquartile range was 10.9% -12.4%.

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- This measure is based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- Very good
- Currently the data used to construct this measure is routinely captured as a par6t of hospitals' bulling processes.

Criterion 4: Usability and Use

4. Usability and 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.

- Measure results are publicly reported through CMS's Hospital Inpatient Quality Reporting (IQR) Program.
- In addition, measure results are incorporated into the calculation of hospital payment rates through CMS's Hospital Value Based Purchasing (HVBP) Program.
- The developer reports that there has been significant progress in 30-day mortality rates for heart failure, noting that the median 30-day RSMR decreased by 0.7 absolute percentage points from 2011-2012 (median RSMR: 11.7%) to 2013-2014 (median RSMR: 11.0%).
- The developers did not identify any unintended consequences during measure development or model testing, but note that they are committed to monitoring this measure's use and assessing potential unintended consequences over time.

Questions for the Committee*:*

- *Are this measure's performance results being used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- It is in use. Publically reported.
- The measure results have been and currently are being publicly reported through CMS' IQR Program, as well as

is being incorporated in the Hospital Value Based Purchasing (HVBP) Program. While the unintended consequences were note identified by the developer, there was a note that showcased that there has been progress in 30-day mortality rates for heart failure, noting that the median 30-day RSMR decreased by .7 absolute percentage form 2011-2012 to 2013-2014.

Criterion 5: Related and Competing Measures
• List any related or competing measures based on harmonization protocol. • Summarize any harmonization efforts, i.e., responses from the developers regarding harmonization. • Briefly summarize next steps according to protocol

Pre-meeting public and member comments

•

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older

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

Date of Submission: Click here to enter a date

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related

behavior.

- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

1a.1.This is a measure of:

Outcome

☒ Health outcome: cost/resource use.

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (e.g. lab value): Click here to name the intermediate outcome

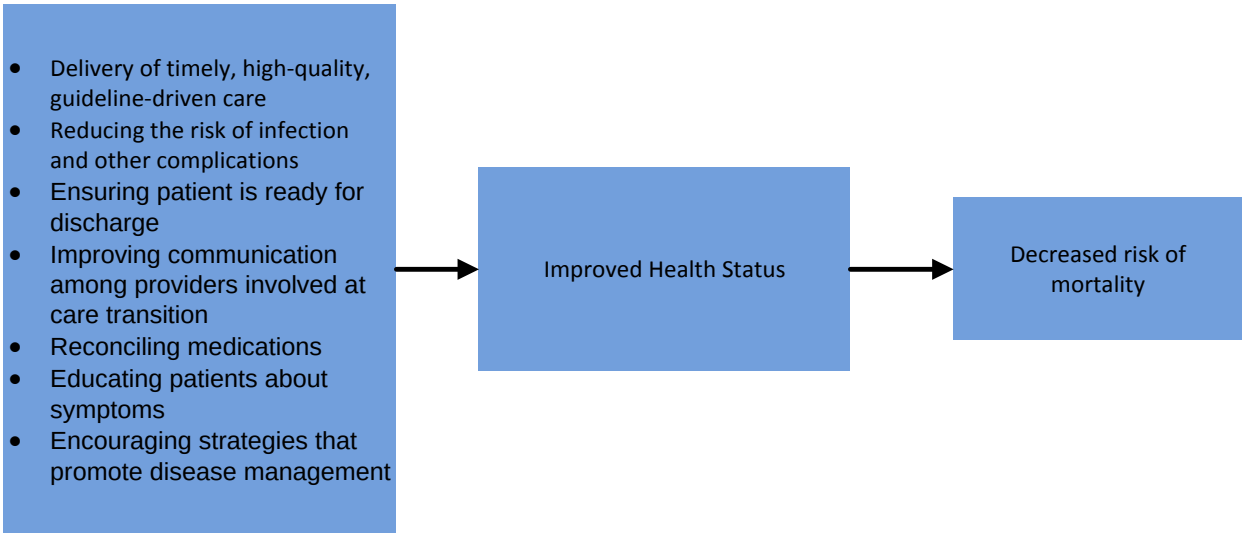
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



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

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Heart failure (HF) incidence approaches 10 per 1000 of the population after 65 years of age (NHLBI 2007), and is the most common discharge diagnosis among the elderly (Jessup and Brozena 2003); prevalence of HF in the U.S. is estimated at nearly 6 million (Mozaffarian 2015, Lloyd-Jones 2009), and is suspected as the leading cause of death in people over age 65.

According to the 2015 AHA update report, one in 9 deaths has HF mentioned on the death certificate. In 2011, HF any-mention mortality was 284,388. HF was the underlying cause in 58,309 of those deaths in 2011 (Mozaffarian 2015, National Center for Health Statistics 2011). There are 870,000 new HF cases annually (Mozaffarian 2015). Survival after HF diagnosis has improved over time, however, the death rate remains high:

≈50% of people diagnosed with HF will die within 5 years (Mozaffarian 2015, Levy et al. 2002, Roger et al. 2004). Among Medicare beneficiaries, the overall 1-year HF mortality rate declined slightly from 1998 to 2008 but remained high at 29.6% (Chen et al. 2011). Rates of mortality decline were uneven across states.

Clinical experience suggests that the care for these patients is highly variable, and studies suggest quality gaps in hospital care—particularly in the transition to outpatient care (Albert 2009, Jha 2005). Moreover, there is substantial inter-hospital variation in the risk of death that is not clearly explained by differences in case mix. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures.

The HF RSMR measure is thus intended to inform quality-of-care improvement efforts, as individual process-based performance measures cannot encompass all the complex and critical aspects of care within a hospital that contribute to patient outcomes. Many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals.

The diagram above indicates some of the many care processes that can influence mortality risk. Numerous studies have demonstrated that appropriate and timely treatment for HF patients can reduce the risk of mortality within 30 days of hospital admission (Hunt 2009, Jha 2007). Additionally, trials of interventions which improve patient education upon discharge have been shown to improve survival for HF patients (McAllister 2001). Evidence that hospitals have been able to reduce mortality rates through these quality-of-care initiatives illustrates the degree to which hospital practices can affect mortality rates.

References

Albert NM, Yancy CW, Liang L, et al. Use of aldosterone antagonists in heart failure. *JAMA*. 2009;302(15):1658-1665.

Chen J, Normand SL, Wang Y, Krumholz HM. National and regional trends in heart failure hospitalization and mortality rates for Medicare beneficiaries, 1998–2008. *JAMA*. 2011;306:1669–1678.

Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, Jessup M,Konstam MA, Mancini DM, Michl K, Oates JA, Rahko PS, Silver MA, Stevenson LW,Yancy CW; American College of Cardiology Foundation; American Heart Association.2009 Focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the International Society for Heart and Lung Transplantation. J Am Coll Cardiol. 2009 Apr 14;53(15):e1-e90.

Jessup M, Brozena S. Medical progress: heart failure. N Engl J Med 2003;348:2007–18.

Jha AK, Li Z, Orav EJ, Epstein AM. Care in U.S. Hospitals — The Hospital Quality Alliance Program. *New England Journal of Medicine.* 2005;353(3):265-274.

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

Levy D, Kenchaiah S, Larson MG, Benjamin EJ, Kupka MJ, Ho KK, Murabito JM, Vasan RS. Long-term trends in the incidence of and survival with heart failure. *N Engl J Med*. 2002;347:1397–1402.

Lloyd-Jones D et al,American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010 Feb 23;121(7):e46-e215. Epub 2009 Dec 17.

McAllister FA, Lawson FME, Teo KK, Armstrong PW: A systematic review of randomized trials of disease management programs in heart failure. Am J Med 2001, 110:378-384

Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation.* 2015;131(4):e29-322.

National Center for Health Statistics. Mortality Multiple Cause Micro-data Files, 2011. Public-use data file and documentation. NHLBI tabulations.
http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm#Mortality_Multiple. Accessed July 3, 2014.

National Heart, Lung, and Blood Institute. Unpublished tabulation of NHANES, 1971-1975, 1976-1980, 1988-1994, 1999-2002, 2003-2006, and extrapolation to the U.S. population, 2007.

Roger VL, Weston SA, Redfield MM, Hellermann-Homan JP, Killian J, Yawn BP, Jacobsen SJ. Trends in heart failure incidence and survival in a community-based population. *JAMA*. 2004;292:344–350.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
HF_Mortality_NQF_Evidence_Attachment_06-29-15.docx

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

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

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Distribution of Hospital HF RSMRs over Different Time Periods
Results for each data year
Characteristic//07/2011-06/2012//07/2012-06/2013//07/2013-06/2014//07-2011-06/2014
Number of Hospitals// 4,671 // 4,651 // 4,597 // 4,775
Number of Admissions// 333,279 // 332,507 // 325,221 // 991,007
Mean (SD)// 11.8 (1.1) // 12.1 (0.9) // 11.1 (0.8) // 11.7 (1.3)
Range (min. – max.)// 7.6-17.9 // 8.3-17.3 // 7.6-15.4 // 7.0-19.3
Minimum// 7.6 // 8.3 // 7.6 // 7.0
10th percentile// 10.6 // 11.1 // 10.1 // 10.1
20th percentile// 11.1 // 11.5 // 10.5 // 10.7
30th percentile// 11.4 // 11.7 // 10.8 // 11.1
40th percentile// 11.6 // 11.9 // 10.9 // 11.4
50th percentile// 11.7 // 12.0 // 11.0 // 11.6
60th percentile// 11.9 // 12.2 // 11.2 // 11.9
70th percentile// 12.2 // 12.4 // 11.4 // 12.3
80th percentile// 12.5 // 12.8 // 11.6 // 12.6
90th percentile// 13.1 // 13.2 // 12.1 // 13.4
Maximum// 17.9 // 17.3 // 15.4 // 19.3

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

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Distribution of HF RSMRs by Proportion of Medicaid Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic//Hospitals with a low proportion (=7.0%) Medicaid patients//Hospitals with a high proportion (=29.6%) Medicaid patients
Number of Measures Entities (Hospitals)// 373 // 373
Number of Patients// 84,068 patients in low-proportion hospitals/74,416 in high-proportion hospitals
Maximum// 15.1 // 16.4
90th percentile// 13.2 // 13.3
75th percentile// 12.4 // 12.4
Median (50th percentile)// 11.6 // 11.3
25th percentile// 10.9 // 10.5
10th percentile// 10.1 // 9.5
Minimum // 7.7 // 7.2

Distribution of RSMRs by Proportion of African-American Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic// Hospitals with a low Proportion (=0%) African-American patients//Hospitals with a high proportion (=23.3%) African-American patients
Number of Measures Entities (Hospitals)// 526 // 376
Number of Patients//31,904 patients in low-proportion hospitals/92,579 in high-proportion hospitals
Maximum// 16.4 // 15.9
90th percentile// 13.8 // 12.9
75th percentile// 12.8 // 12.0
Median (50%)// 12.0 // 11.1
25th percentile// 11.3 // 10.2
10th percentile// 10.8 // 9.3
Minimum// 9.1 // 7.2

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
N/A

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Severity of illness, Patient/societal consequences of poor quality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
HF incidence approaches 10 per 1000 population after 65 years of age (NHLBI 2007), and is the most common discharge diagnosis among the elderly (Jessup and Brozena 2003); prevalence of HF in the U.S. is estimated at nearly 6 million. (Lloyd-Jones 2009), and is suspected to be the leading cause of death in people over age 65.

Many current hospital interventions are known to decrease the risk of death within 30 days of hospital admission (Jha 2007). Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative hospital performance on outcomes measures.

Numerous studies have demonstrated that appropriate and timely treatment for HF patients can reduce the risk of mortality within 30 days of hospital admission. (Hunt 2009, Jha 2007) Additionally, trials of interventions which improve patient education upon discharge have been shown to improve survival for HF patients (Mcalister 2001).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, Jessup M,Konstam MA, Mancini DM, Michl K, Oates JA, Rahko PS, Silver MA, Stevenson LW,Yancy CW; American College of Cardiology Foundation; American Heart Association.2009 Focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the International Society for Heart and Lung Transplantation. J Am Coll Cardiol. 2009 Apr 14;53(15):e1-e90.

1a.4. Citations for Evidence of High Impact

Jessup M, Brozena S. Medical progress: heart failure. N Engl J Med 2003;348:2007–18.

National Heart, Lung, and Blood Institute. Unpublished tabulation of NHANES, 1971-1975, 1976-1980, 1988-1994, 1999-2002, 2003-2006, and extrapolation to the U.S. population, 2007.

Lloyd-Jones D et al,American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010 Feb 23;121(7):e46-e215. Epub 2009 Dec 17

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

McAllister FA, Lawson FME, Teo KK, Armstrong PW: A systematic review of randomized trials of disease management programs in heart failure. Am J Med 2001, 110:378-384

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*
Safety

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.qualitynet.org/dcs/ContentServer?cid=1163010421830&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page & http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** HF_Mortality_NQF_Data_Dictionary__06-22-15_FINAL.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Annual Updates
1. Updated CC map.
a. Rationale: The ICD-9-CM CC map was updated annually to capture all relevant comorbidities coded in patient administrative claims data.

No other updates made after 2013 Measure Updates except for use of new years of data for public reporting

Planned Update for 2016 public reporting – (changes reflected in this application)
1. Exclude patients undergoing LVAD implantation or heart transplantation during an index admission or who have a history of LVAD or heart transplant in the preceding year.
a. Rationale: Patients undergoing implantation of an LVAD designed to offer intermediate to long-term support (weeks to years) as a bridge to heart transplant or destination therapy represent a clinically distinct, highly-selected group of patients cared for at highly specialized medical centers. This is a subgroup of patients that has grown in recent years and that have distinct mortality risk and risk factors from the primary heart failure cohort.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days of the index admission date for patients 18 and older discharged from the hospital with a principal diagnosis of HF.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator time window: We define the time period for death from any cause within 30 days from the date of admission for the index HF hospitalization.

Denominator time window: This measure was developed with 12 months of data. The time window can be specified from one to three years. Currently, the measure is publicly reported with three years of index hospitalizations.

S.6. **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, 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.
The measure counts deaths for any cause within 30 days of the date of admission of the index HF hospitalization.

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

Identifying deaths in the all-payer measure
For the purposes of development of an all-payer measure, deaths were identified using the California vital statistics data file. Nationally, post-discharge deaths can be identified using an external source of vital status, such as the Social Security Administration's Death Master File (DMF) or the Centers for Disease Control and Prevention's National Death Index (NDI).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 18 years or older. We have explicitly tested the measure in both age groups. The cohort includes admissions for patients aged 18 years and older discharged from the hospital with a principal discharge diagnosis of HF and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients 65 years and older who are either Medicare FFS beneficiaries admitted to non-federal hospitals or patients admitted to VA hospitals. Additional details are provided in S.9 Denominator Details.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
To be included in the measure cohort used in public reporting, patients must meet the following additional inclusion criteria:
1. Have a principal discharge diagnosis of heart failure
2. Enrolled in Medicare fee-for-service (FFS)
3. Aged 65 or over
4. Discharged from non-federal acute care hospitals or VA hospitals
5. Enrolled in Part A and Part B Medicare for the 12 months prior to the date of index admission.

VA beneficiaries/hospitalizations are also included in the HF mortality measure. Enrollment in Medicare FFS is not required for these patients.

This measure can also be used for an all-payer population aged 18 years and older. We have explicitly tested the measure in both patients aged 18+ years and those aged 65+ years (see Testing Attachment for details).

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for each measure are:
402.01 Malignant hypertensive heart disease with heart failure
402.11 Benign hypertensive heart disease with heart failure
402.91 Unspecified hypertensive heart disease with heart failure
404.01 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.03 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease
404.11 Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.13 Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease stage V or end stage renal disease
404.91 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.93 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease
428.0 Congestive heart failure, unspecified

428.1 Left heart failure
428.20 Systolic heart failure, unspecified
428.21 Acute systolic heart failure
428.22 Chronic systolic heart failure
428.23 Acute on chronic systolic heart failure
428.30 Diastolic heart failure, unspecified
428.31 Acute diastolic heart failure
428.32 Chronic diastolic heart failure
428.33 Acute on chronic diastolic heart failure
428.40 Combined systolic and diastolic heart failure, unspecified
428.41 Acute combined systolic and diastolic heart failure
428.42 Chronic combined systolic and diastolic heart failure
428.43 Acute on chronic combined systolic and diastolic heart failure
428.9 Heart failure, unspecified

ICD-10 Codes that define the patient cohort:
I110 Hypertensive heart disease with heart failure
I130 Hypertensive heart and chronic kidney disease with heart failure and stage 1 through stage 4 chronic kidney disease, or unspecified chronic kidney disease
I132 Hypertensive heart and chronic kidney disease with heart failure and with stage 5 chronic kidney disease, or end stage renal disease
I509 Heart failure, unspecified
I501 Left ventricular failure
I5020 Unspecified systolic (congestive) heart failure
I5021 Acute systolic (congestive) heart failure
I5022 Chronic systolic (congestive) heart failure
I5023 Acute on chronic systolic (congestive) heart failure
I5030 Unspecified diastolic (congestive) heart failure
I5031 Acute diastolic (congestive) heart failure
I5032 Chronic diastolic (congestive) heart failure
I5033 Acute on chronic diastolic (congestive) heart failure
I5040 Unspecified combined systolic (congestive) and diastolic (congestive) heart failure
I5041 Acute combined systolic (congestive) and diastolic (congestive) heart failure
I5042 Chronic combined systolic (congestive) and diastolic (congestive) heart failure
I5043 Acute on chronic combined systolic (congestive) and diastolic (congestive) heart failure

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
The mortality measures exclude index admissions for patients:
1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility.
2. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission;
4. Discharged against medical advice (AMA); or
5. Patients undergoing LVAD implantation or heart transplantation during an index admission or who have a history of LVAD or heart transplant in the preceding year.

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

For Medicare FFS patients, the measure additionally excludes admissions for patients without at least 30 days post-discharge enrollment in FFS Medicare (because the 30-day mortality outcome cannot be assessed in this group).

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

1. The discharge disposition indicator is used to identify patients alive at discharge. Transfers are identified in the claims when a patient with a qualifying admission is discharged from an acute care hospital and admitted to another acute care hospital on the same day or next day. Patient length of stay and condition is identified from the admission claim.
2. Inconsistent vital status or unreliable data are identified if any of the following conditions are met 1) the patient's age is greater than 115 years: 2) if the discharge date for a hospitalization is before the admission date; 3) if the patient has a sex other than 'male' or 'female'.
3. Hospice enrollment in the 12 months prior to or on the index admission is identified using hospice data and the Inpatient standard analytic file (SAF). This exclusion applies when the measure is used in Medicare FFS patients only.
4. Discharges against medical advice (AMA) are identified using the discharge disposition indicator.
5. Patients with LVAD implantation or heart transplantation during an index admission or in the previous 12 months are identified by the corresponding codes for these procedures included in claims data.

Additional exclusions:
• HF admissions within 30 days of discharge from a qualifying index admission, which are identified by comparing the discharge date from the index admission with the readmission date.
• Admissions without at least 30 days post-discharge enrollment in FFS Medicare are determined by examining the Medicare Enrollment Database (EDB)

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

The measure employs a hierarchical logistic regression model to create a hospital-level 30-day RSMR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, the model adjusts the log-odds of mortality within 30 days of admission for age, sex, and selected clinical covariates. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a death at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables: Candidate variables were patient-level risk-adjustors that were expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment, including age, sex, and indicators of comorbidity and disease severity. For each patient, covariates are obtained from claims records extending 12 months prior to and including the index admission. For the measure currently implemented by CMS, these risk-adjusters are identified using both inpatient and outpatient Medicare FFS claims data. However, in the all-payer hospital discharge database measure, the risk-adjustment variables can be obtained only from inpatient claims in the prior 12 months and the index admission. (This was tested explicitly in our all-payer testing, as many all-payer datasets do not include outpatient claims.)

The model adjusts for case-mix differences based on the clinical status of patients at the time of admission. We use condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care and that are only recorded in the index admission.

The final set of risk adjustment variables is:

Demographics
Male
Age-65 (years above 65, continuous) for 65 and over cohorts; or Age (years, continuous) for 18 and over cohorts.

Comorbidities
Congestive heart failure (CC 80)
Acute myocardial infarction (CC 81)
Other acute/subacute forms of ischemic heart disease (CC 82)
Coronary atherosclerosis or angina (CC 83-84)
Cardio-respiratory failure and shock (CC 79)
Valvular and rheumatic heart disease (CC 86)
Hypertension (CC 89, 91)
Stroke (CC 95-96)
Renal failure (CC 131)
Chronic obstructive pulmonary disease (COPD) (CC 108)
Pneumonia (CC 111-113)
Diabetes mellitus (DM) or DM complications except proliferative retinopathy (CC 15-20, 120)
Protein-calorie malnutrition (CC 21)
Dementia or other specified brain disorders (CC 49-50)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)
Vascular disease and complications (CC 104-105)
Metastatic cancer, acute leukemia and other severe cancers (CC 7-8)
Trauma in last year (CC 154-156, 158-162)
Major psychiatric disorders (CC 54-56)
Chronic Liver Disease (CC 25-27)
History of CABG (ICD-9-CM V45.81, 36.10-36.16)
History of PTCA (ICD-9-CM V45.82, 00.66, 36.01, 36.02, 36.05, 36.06, 36.07)
References:
Krumholz HM, Brindis RG, Brush JE, et al. 2006. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation 113: 456-462.

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

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Available at measure-specific web page URL identified in S.1.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*

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

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

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

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed readmission rate. The hierarchical logistic regression models are described fully in the original methodology report [Krumholz et al., 20052].

Reference:

1. Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.
2. Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Models for AMI and HF 30-Day Mortality Methodology. 2005.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

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

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF a PRO-PM, identify whether (and how) proxy responses are allowed.

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

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*

IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

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

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)

Required for Composites and PRO-PMs.

Missing values are rare among variables used from claims data in this measure.

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.

Administrative claims, Other, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Data sources for the Medicare FFS measure:
1. Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician claims for the 12 months prior to an index admission.
2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992).
3. Veterans Health Administration (VA) Data: This data source contains claims data for VA inpatient and outpatient services including: inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician claims for the 12 months prior to and including each index admission. Unlike Medicare FFS patients, VA patients are not required to have been enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission.

All-payer data sources:
For our analyses to examine use in all-payer data, we used all-payer data from California in addition to CMS data for Medicare FFS 65+ patients in California hospitals. California is a diverse state, and, with more than 37 million residents, California represents 12% of the US population. We used the California Patient Discharge Data, a large, linked database of patient hospital admissions. In 2006, there were approximately 3 million adult discharges from more than 450 non-Federal acute care hospitals. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations and to evaluate rates of both readmission and mortality (via linking with California vital statistics records).

Using all-payer data from California as well as CMS Medicare FFS data for California hospitals, we performed analyses to determine whether the HF mortality measure can be applied to all adult patients, including not only FFS Medicare patients aged 65+ but also non-FFS Medicare patients aged 18-64 years at the time of admission.

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

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
HF_Mortality_new_testing_attachment_v1_1.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0229

Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older

Date of Submission:[enter submission date]

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☒ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/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: Census Data/American Community Survey

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

The datasets used for testing included Medicare Parts A and B claims, Veterans' Health Administration claims, as well as the Medicare Enrollment Database (EDB). Additionally, census data were used to assess sociodemographic factors. The dataset used varies by testing type; see Section 1.7 for details.

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

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

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**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*)

For this measure, hospitals are the measured entities. All non-federal, acute inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries aged 65 years and older are included. All Veteran's Health Administration Hospitals are also included in the current publically reported measure. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

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

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

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

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

For reliability testing (Section 2a2)
The reliability of the model was tested by randomly selecting 50% of the Medicare patients aged 65+ in the most recent 3-year cohort and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients and compared the two. Thus, for reliability testing, we randomly split Dataset 1 into two samples. In each year of measure maintenance, we also re-fit the model and compare the frequencies and model coefficients of risk variables (condition categories for patient comorbidities) and model fit across 3 years (Dataset 1 below).

Dataset 1 (current public reporting cohort): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: July 1, 2011 – June 30, 2014 (current public reporting cohort)
Number of Admissions: 991,007
Patient Descriptive Characteristics: average age=81.1, % male=46.3084
Number of Measured Entities: 4,775

For validity testing (Section 2b2)
Dataset 2 (medical record validation): Chart Validation: National Heart Failure (NHF) Dataset for clinical data from HF hospital admissions, linked with the Medicare Part A Inpatient and Outpatient and Part B Outpatient claims and the Medicare Enrollment Database to assess the mortality outcome.
Dates of Data: 1998-2001
Number of Admissions: N=46,700

For testing of measure exclusions (Section 2b3)
Dataset 1 (current public reporting cohort)

For testing of measure risk adjustment (Section 2b4)
Dataset 1 (current public reporting cohort)

Dataset 3 (development dataset): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: 1998
Number of Admissions: N=222,424 (first half of split sample); N=222,157 (second half of split sample)
Number of Measured Entities: 5,087 (first half of split sample); 5,088 (second half of split sample)

Dataset 4 (all payer dataset): California Patient Discharge Data in addition to CMS Medicare FFS data for patients in California hospitals
Dates of Data: January 1, 2006 – December 31, 2006
Number of Discharges: 60,022 (all 18+ total); 27,977 (FFS 65+); 16,447 (non-FFS 65+); 15,598 (all 18-64)
Patient Descriptive Characteristics: mean age=73, %male=49 (all 18+ total); mean age=81, %male=43 (FFS 65+); mean age=80, %male=46 (non-FFS 65+); mean age=53, %male=61 (all 18-64)
Number of Measured Entities: >450 non-Federal acute care hospitals
The measure was applied to California Patient Discharge Data, a large, linked all-payer database of patient hospital admissions. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations. In addition, the unique patient ID number is used to link with state vital statistics records to assess 30-day mortality.

For testing to identify meaningful differences in performance (Section 2b5)
Dataset 1 (current public reporting cohort)

For testing of socio-demographic factors in risk models (Section 2b4.3)
Dataset 1 (current public reporting cohort)

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We selected sociodemographic status (SDS) variables to analyze after reviewing the literature and examining available national data sources. Few patient-level SDS variables that can be linked to Medicare data are available nationally. We found both race [black vs non-black] and dual-eligible status, e.g. enrolled in both Medicare and Medicaid, [obtained from CMS claims enrollment data] as the only two patient-level SDS variables available to examine directly. While some argue against consideration of race in risk-adjustment, we felt it was important to understand the association with race as well as more traditional socio-economic variables.

We also considered neighborhood-level variables, linked by patient zip code, that could serve in a risk model as a proxy for patient-level SDS. A range of census-collected SDS variables [collected annually as part of American Community Survey and aggregated over 5-years] including income and education, were available. Currently, we are only able to link the data at a 5-digit zip code level. Nine-digit zip code data may provide a more granular view of patient sociodemographic status, but this data is not available to us at this time; we therefore cannot ascertain the incremental, if any, value of greater geographic discrimination for risk adjustment purposes.

Our conceptual model and the literature regarding how SDS may influence post-discharge mortality did not identify a single SDS concept as predominant in the pathway. There is a large body of literature linking various SDS factors to worse health status and higher mortality over a lifetime (Adler and Newman 2002, Mackenbach et al. 2000). Income, education, and occupational level are the most commonly examined variables. However, literature directly examining how different SDS factors might influence the likelihood of older, insured, Medicare patients of dying within 30 days of an admission for cardiovascular disease is much more limited. Assuming that the risk imparted based on zip code may reflect multiple different SDS variables, we chose to analyze a validated AHRQ composite index of SDS which has been used and tested among Medicare beneficiaries (Blum et al. 2014; Bonito et al. 2008). This index is a composite of 7 different variables found in the census data which may capture SDS better than any single variable. The index variables include rates of

unemployment, percent of person living below poverty, education level (percent below 12th grade education and percent with college education), crowding (average of more than one person per room) median household income and median housing value. We identified patients as low SDS if they lived in a neighborhood in lowest quartile of this index.

Other variables can be found at a county or regional level and could represent the hospital's community. We did not directly test any such variables because they are not as closely related to patient's sociodemographic status given the wide scope of a county and seemed unlikely to be ideal for patient-level risk adjustment.

References:

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

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

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

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

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*)

Data Element Reliability

In constructing the measure, we aim to utilize only those data elements from the claims that have both face validity and reliability. We avoid the use of fields that are thought to be coded inconsistently across hospitals or providers. Specifically, we use fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of CMS auditing and billing policies and seek to avoid variables which do not meet this standard. For example, "discharge disposition" is a variable in Medicare claims data that is not thought to be a reliable variable for identifying a transfer between two acute care facilities. Thus, we derive a variable using admission and discharge dates as a surrogate for "discharge disposition" to identify hospital admissions involving transfers. This allows us to identify these admissions using variables in the claims data which have greater reliability than the "discharge disposition" variable. In addition, CMS has in place several hospital auditing programs used to assess overall claims code accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in our measures, including diagnosis and procedure codes and other elements that are consequential to payment.

Finally, we assess the reliability of the data elements by comparing model variable frequencies and odds ratios from logistic regression models across three years of data.

Measure Score reliability

The reliability of a measurement is the degree to which repeated measurements of the same entity agree with each other. For measures of hospital performance, the measured entity is naturally the hospital, and reliability is the extent to which repeated measurements of the same hospital give similar results. In line with this thinking, our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients produces similar measures of hospital performance. That is, we take a "test-retest" approach in which hospital performance is measured once using a random subset of patients, then measured again using a second random subset exclusive of the first, and finally comparing the agreement between the two resulting performance measures across hospitals (Rousson et al., 2002). For test-retest reliability, we combined index admissions from successive measurement periods into one dataset, randomly sampled half of patients within each hospital, calculated the measure for each hospital, and repeated the calculation using the second half. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement we calculated the intra-class correlation coefficient (ICC) (Shrout and Fleiss, 1979), and assessed the values according to conventional standards (Landis and Koch, 1977). Specifically, we used dataset 1 split sample and calculated the RSRR for each hospital for each sample. The agreement of the two RSRRs was quantified for hospitals using the intra-class correlation as defined by ICC (2,1) by Shrout and Fleiss (1979).

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

References:

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

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-428.

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

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

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

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

Data element reliability results (Dataset 1)

The frequency of some model variables increased between 2011 and 2014, which may reflect an increased rate of comorbidity in the FFS population, but may also be due, in part, to increased coding opportunities on administrative claims. In the 2012 update to the measures, we increased the number of diagnosis and procedure codes to align with the version 5010 format changes DHHS required. Hospitals could begin to submit up to 25 diagnosis and procedure codes starting in 2010. Over time, more hospitals submitted more codes, which translated into increased frequencies for some model variables. Notable decreases occurred in Coronary atherosclerosis or angina (CC 83-84) (73.5% to 72.0%), Congestive heart failure (CC 80) (74.8% to 73.7%), and Acute myocardial infarction (CC 81) (9.7% to 9.5%). Notable increases occurred in Cardio-respiratory

failure or shock (CC 79) (25.6% to 27.9%), History of Percutaneous Transluminal Coronary Angioplasty (PTCA) (ICD-9 codes V45.82, 00.66, 36.06, and 36.07) (12.7% to 13.8%), and Hypertension (CC 89, 91) (93.3% to 93.8%). Examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across three years.

These frequencies are from the model containing the LVAD exclusion, which will be incorporated into the Annual Updates Reports in 2016. For the model variable frequencies and risk variable odds ratios without the LVAD exclusion, see the 2015 Annual Updates Report (Dorsey et al. 2015).

Measure Score Reliability Results (Dataset 1)
There were 991,007 admissions in the combined 3-year sample, with 494,297 in one sample and 496,710 in the other randomly selected sample. The agreement between the two RSMRs for each hospital was 0.55, which according to the conventional interpretation is "moderate" (Landis & Koch, 1977).
Note that this analysis was limited to hospitals with 12 or more cases in each split sample. The intra-class correlation coefficient is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is reported with the full three years of data. The correlation coefficient is expected to be higher using the full three-year sample since it would include more patients.

Reference:

Dorsey K, Grady J, Desai N, et al. 2015 Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Mortality Measures Acute Myocardial Infarction – Version 9.0 Heart Failure – Version 9.0 Pneumonia – Version 9.0 Chronic Obstructive Pulmonary Disease – Version 4.0 Stroke – Version 4.0. 2015; https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890435185&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DMort_AMI-HF-PN_MsrUpdRpt_32715.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed June 15, 2015.

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

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

The stability over time of the risk factor frequencies and odds ratios suggests that the underlying data elements are reliable. Additionally, the ICC score demonstrates moderate agreement across samples using a conservative approach to assessment.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

During original measure development we validated the HF mortality administrative model against a medical record model in the same cohort of patients for which hospital-level HF mortality medical record data are available.

We developed a medical record measure to compare with the administrative measure. We developed a measure cohort with the medical record data using the inclusion/exclusion criteria and risk-adjustment strategy that was consistent with the claims-based administrative measure but using chart-based risk adjusters, such as blood pressure, not available in the claims data. We then matched a sample of the same patients in the administrative data for comparison. The matched sample included 46,700 patients. We compared the output of the two measures, the state performance results, in the same group of patients.

For the derivation of the chart-based model, we used cases identified through a Health Care Financing Administration (now CMS) quality initiative, which sampled admissions from FFS Medicare beneficiaries for several clinical conditions, including HF (Jencks et al., 2000). Cases were identified over a 6-month period within each state, plus the District of Columbia and Puerto Rico, during the period April 1, 1998 through October 31, 1999. Based on the principal discharge diagnosis, approximately 800 HF discharges per state were identified, and the corresponding medical records were abstracted by 2 clinical data abstraction centers. In states with fewer than 900 HF discharges, all cases were used. The abstractors first sorted eligible claims by age, race, sex, and hospital. Then, they systematically sampled cases from a random starting point. Patients must have been enrolled in FFS Medicare. CMS subsequently conducted a re-measurement using the same data collection methodology for 2000 and 2001 discharges (Jencks et al., 2003), and the combined 1998-2001 data, including 73,832 patients, served as the NHF dataset for development of the chart-based model.

From the medical chart-abstracted HF cases, we linked these files to the corresponding administrative data and mortality data from the Medicare enrollment database. Because only patients aged 65 years and older were included, and some data were excluded based on linkage and other factors, a total of 46,700 HF hospitalizations were used in the analysis**.**

The same coding and transfer rules described in the HF administrative dataset were used in defining the HF chart dataset.

The chart model was derived in the NHF dataset. The derivation sample contained 46,700 cases with an unadjusted 30-day mortality rate of 11.9%. Twenty-eight covariates were included in the final model, with age having the largest impact on risk. While the administrative mortality models explained about 10-12% of the observed variation and had accuracy of 69-71%, the chart model explained 21-22% of the variation and had accuracy of 75-78%. Moreover, the predictive ability of the model is excellent—observed mortality in the lowest estimated decile is 1.8% for 30-day mortality, compared with 42.4% (30-day mortality) in the highest estimated decile, a range of 40.6%.

<u>Validity Indicated by Established Measure Development Guidelines:</u>
We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes measurement set forth in NQF guidance for outcomes measures (National Quality Forum, 2010), CMS Measure Management System (MMS) guidance, and the guidance articulated in the American Heart Association scientific statement, “Standards for Statistical Models Used for Public Reporting of Health Outcomes” (Krumholz, Brindis, et al. 2006).

Citations:

Jencks SF, Cuerdon T, Burwen DR et al. Quality of medical care delivered to Medicare beneficiaries: a profile at state and national levels. JAMA. 2000;284:1670-1676.

Jencks SF, Huff ED, Cuerdon T. Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001. JAMA. 2003;289:305-312.

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

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

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

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

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

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

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

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were initially identified using 2015 GEM mapping software. We then enlisted the help of clinicians with expertise in relevant areas to select and evaluate which ICD-10 codes map to the ICD-9 codes currently in use for this measure. We verify each year that there are no changes. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The performance of the administrative and medical record models is similar (Dataset 2). The areas under the receiver operating characteristic (ROC) curve for the two models are 0.71 and 0.78, respectively.

We estimated hospital-level RSMRs using the corresponding hierarchical logistic regression administrative and medical record models for the linked patient sample. We then examined the linear relationship between the two sets of estimates using regression techniques and weighting by the total number of cases in each hospital. The correlation coefficient of the standardized rates from the administrative and medical record models is 0.95.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The results between the administrative and medical record models proved to be similar in each of the model testing that was performed. The ROC results were nearly identical and in line with other mortality models. The correlation between the resulting RSMRs calculated from both models was 0.95, which shows the resulting measure from the administrative claims model is as good as that from the medical record model.

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (Dataset 1). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

2b3.2. 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*)

Among 4,775 hospitals with at least 25 index stays in July 1, 2011 – June 30, 2014:

Exclusion	N	%	Distribution across hospitals: Min, 25th, 50th, 75th percentile, max
1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility	88,023	6.45%	(0, 3.51, 6.07, 9.33, 100.00*) *due to small size hospital

2. Inconsistent or unknown vital status or other unreliable demographic (age and gender) data	55	<0.01%	(0, 0, 0, 0, 2.86)
3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission	18,753	1.37%	(0, 0, 0.88, 1.96, 50.00)
4. LVAD or Transplant in index admission or prior year	2,362	0.17%	(0, 0, 0, 0.49, 50.00)
5. Discharged against medical advice (AMA)	5,933	0.43%	(0, 0, 0, 0, 11.52)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusion 1 (patients who were discharged alive on the day of admission or the following day who were not transferred to another acute care facility) accounts for 6.45% of all index admissions excluded from the initial index cohort, the majority of all exclusions, and is meant to ensure a clinically coherent cohort. This exclusion prevents inclusion of patients who likely did not have clinically significant HF. One of the remaining four exclusions applies to less than 2% of admissions, and the other three apply to less than 1% of admissions. For exclusion 2 (patients with inconsistent or unknown vital status or other unreliable demographic (age and gender) data), we do not include stays for patients where the age is greater than 115, where the gender is neither male nor female, where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge but the patient was discharged alive. For exclusion 3 (patients enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission), these patients are likely continuing to seek comfort measures only; thus, mortality is not necessarily an adverse outcome or signal of poor quality care. Exclusions 4 (patients with LVAD, history of LVAD, transplant, history of transplant) are meant to ensure a clinically coherent cohort. Patients undergoing implantation of an LVAD designed to offer intermediate to long-term support (weeks to years) as a bridge to heart transplant or destination therapy represent a clinically distinct, highly-selected group of patients cared for at highly specialized medical centers. Exclusion 5 (patients who are discharged AMA) is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge.

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

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 24 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Our approach to risk adjustment was tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al. 2006).

The measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day RSMR. This approach to modeling appropriately accounts for the structure of the data (patients clustered within hospitals), the underlying risk due to patients' comorbidities, and sample size at a given hospital when estimating hospital mortality rates. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand and Shahian et al. 2007). At the patient level, each model adjusts the log-odds of mortality within 30-days of admission for age, sex, selected clinical covariates and a hospital-specific intercept. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept, or hospital-specific effect, represents the hospital contribution to the risk of mortality, after accounting for patient risk and sample size, and can be inferred as a measure of quality. The hospital-specific intercepts are given a distribution in order to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Clinical Factors
Candidate and Final Risk-adjustment Variables: The original measure was developed using Medicare FFS claims data. Candidate variables were patient-level risk-adjustors that are expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity. For each patient, covariates were obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusted for case differences based on the clinical status of the patient at the time of admission. We used condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes. We did not risk-adjust for CCs that were possible adverse events of care and that were only recorded in the index admission. In addition, only comorbidities that conveyed information about the patient at that time or in the 12-months prior, and not complications that arose during the course of the hospitalization were included in the risk-adjustment.

The final set of risk-adjustment variables is:

Demographic
• Age-65 (years above 65, continuous) for 65 and over cohorts; or Age (years, continuous) for 18 and over cohorts
• Male

Cardiovascular
• History of PTCA
• History of CABG
• Congestive heart failure
• Acute myocardial infarction
• Unstable angina
• Chronic atherosclerosis
• Cardio-respiratory failure and shock
• Valvular and rheumatic heart disease

Comorbidity
• Hypertension
• Stroke
• Renal failure
• Pneumonia
• Diabetes and DM complications
• Protein-calorie malnutrition
• Dementia and senility
• Hemiplegia, paraplegia, paralysis, functional disability
• Peripheral vascular disease
• Metastatic cancer, acute leukemia, and other severe cancers
• Trauma in last year
• Major psych disorders
• Chronic liver disease

Sociodemographic Factors

We selected candidate sociodemographic factors for examination based on a review of literature, conceptual pathways and feasibility. In section 1.8, we describe the variables that we considered and analyzed based on this review. Below we describe the pathways by which SDS may influence 30-day mortality.

Our conceptualization of the pathways by which patient SDS affects 30-day mortality is informed by the literature. However, as noted previously, although there is a long list of studies showing a relationship between lower SDS status and mortality rates generally, there is relatively little literature directly examining how SDS might influence the likelihood of older, insured, Medicare patients dying within 30 days of a admission for cardiovascular disease and even less literature to directly illuminate the pathways by which SDS influences this outcome. The ability to distinguish between these pathways is challenging but important for making decisions regarding risk adjustment.

One important pathway by which patients' SDS influences 30-day mortality is through health status at the time of admission. SDS factors can influence admission health status both due to the impact of multiple related stressors over a lifetime contributing to overall worse health as well as through poorer access to care and potentially delayed presentation. This results in low SDS patients, when compared with other patients, often arriving for hospital admission with greater levels of illness or comorbidity burden. This pathway should be largely accounted for in our current clinical risk-adjustment model.

However, there are a number of other pathways by which patient SDS may influence 30-day mortality that are related to hospital quality. The first, sometimes referred to as contextual effects, is the instance of patients of low SDS that may present at lower quality institutions for care. Therefore, some part of the apparent relationship between SDS and mortality may be due to clustering of patients of low SDS at poorer quality institutions (Barnato et al. 2005; Hasnain-Wynia et al. 2010; Jha et al. 2011; Skinner et al. 2005).

Next, within the hospital, patients of low SDS may receive differentiated care as compared to counterparts of higher SDS. This can occur for a variety of reasons, and in some cases, differentiated care could be worse quality and in others better quality. For example, providers may be less likely to offer guideline-concordant care to patients of low SDS – perhaps based on discrimination or misunderstanding of patients' wishes and values (Institute of Medicine, 2009). However, in other cases, differentiated care may be patients of low SDS appropriately needing different types of care or services – such as low literacy information, social worker support, or transportation at discharge. Providing needed differentiated care is patient-centered and appropriate, the equivalent to ensuring a diabetic goes home with insulin. However, low SDS patients may not always receive needed differentiated care. This lack of needed differentiated care may also contribute to relatively worse outcomes.

Finally, there may be pathways by which SDS influences 30-day mortality risk outside of healthcare quality and admission health status. Some SDS factors may affect the likelihood of mortality without directly affecting health status on admission or the quality of care received during the hospital stay. For instance, despite a hospital making appropriate care decisions and providing tailored care and education, a lower-income patient may elect not to follow prescribed care (for example, refill a prescription or keep a follow-up visit with a primary care provider) because limited resources create competing priorities for the patient.

These set of proposed pathways are complex to distinguish analytically. We, therefore, first assessed if there was evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways. First we evaluated the variation in the prevalence of low SDS patients among providers. We then assessed the relationship between the SDS variables and the outcome and examined the incremental effect of SDS in a multivariable model. For these measures, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results. Given no meaningful improvement in the risk-model or change in performance scores we did not further seek to distinguish the causal pathways for these measures.

Based on this model and the considerations outlines in 1.8, the following SDS variables were considered:

- Dual eligible status
- African American race
- AHRQ SES index

References:

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

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

IOM (Institute of Medicine). 2009. Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. Washington, DC: The National Academies Press.

Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and medicaid patients. Health affairs (Project Hope). 2011;30(10):1904-1911.

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

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

Skinner J, Chandra A, Staiger D, Lee J, McClellan M. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation. 2005;112(17):2634-2641.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Below are tables showing the final variables that made it into the model.

Final Model Variables (variables meeting criteria in field 2b4.3) (Dataset 1)

Variable	07/2011-06/2014 OR (95% CI)
Age minus 65 (years above 65, continuous)	1.05 (1.05 - 1.05)
Male	1.29 (1.28 - 1.31)
History of Percutaneous Transluminal Coronary Angioplasty (PTCA) (ICD-9 codes V45.82, 00.66, 36.06, 36.07)	0.74 (0.73 - 0.76)
History of Coronary Artery Bypass Graft (CABG) surgery (ICD-9 codes V45.81, 36.10-36.16)	0.89 (0.87 - 0.90)
Congestive heart failure (CC 80)	1.22 (1.20 - 1.24)
Acute myocardial infarction (CC 81)	1.25 (1.22 - 1.28)
Other acute/subacute forms of ischemic heart disease (CC 82)	0.97 (0.95 - 0.99)
Coronary atherosclerosis or angina (CC 83-84)	0.97 (0.95 - 0.98)
Cardio-respiratory failure or shock (CC 79)	1.18 (1.16 - 1.20)
Valvular or rheumatic heart disease (CC 86)	1.08 (1.06 - 1.09)
Hypertension (CC 89, 91)	0.66 (0.65 - 0.68)
Stroke (CC 95-96)	0.96 (0.94 - 0.98)
Renal failure (CC 131)	1.22 (1.20 - 1.24)
Chronic obstructive pulmonary disease (COPD) (CC 108)	1.06 (1.05 - 1.08)
Pneumonia (CC 111-113)	1.32 (1.31 - 1.34)
Diabetes mellitus (DM) or DM complications except proliferative retinopathy (CC 15-20, 120)	0.98 (0.96 - 0.99)
Protein-calorie malnutrition (CC 21)	1.95 (1.91 - 1.98)
Dementia or other specified brain disorders (CC 49-50)	1.37 (1.35 - 1.39)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177-178)	1.12 (1.09 - 1.14)
Vascular disease and complications (CC 104-105)	1.01 (1.00 - 1.03)
Metastatic cancer, acute leukemia and other severe cancers (CC 7-8)	1.79 (1.74 - 1.83)
Trauma in last year (CC 154-156, 158-162)	1.08 (1.07 - 1.10)
Major psychiatric disorders (CC 54-56)	1.12 (1.10 - 1.14)
Chronic liver disease (CC 25-27)	1.55 (1.50 - 1.60)

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Variation in prevalence of the factor across measured entities

The prevalence of SDS factors in the HF cohort varies across measured entities. The median percentage of dual eligible patients is 13.2% (interquartile range [IQR] 8.2%-20.8%). The median percentage of black patients is 3.3% (IQR 0.0%-12.6%). The median percentage of low SES AHRQ indicator patients is 17.6% (IQR 4.0%-49.7%).

Empirical association with the outcome (univariate)

The patient-level observed HF mortality rate is lower for dual-eligible patients, 10.84%, compared with 11.77% for all other patients. The mortality rate for black patients was also lower at 7.39% compared with 12.19% for patients of all other races. Similarly the mortality rate for patients in the lowest SES quartile by AHRQ Index was 10.85% compared with 11.96% for patients in the highest SES quartile.

Incremental effect of SDS variables in a multivariable model

We then examined the strength and significance of the SDS variables in the context of a multivariable model. Consistent with the above findings, when we include any of these variables in a multivariate model that includes all of the claims-based clinical variables the effect size of each of these variables is small and protective. We also find that the c-statistic is essentially unchanged with the addition of any of these variables into the model. Furthermore we find that the addition of any of these variables into the model has little to no effect on hospital performance. We examined the change in hospitals' RSMRs with the addition of any of these variables. The mean absolute change in hospitals' RSMRs when adding a dual eligibility indicator is -0.00009% with a correlation coefficient between RSMRs for each hospital with and without dual eligibility added of 0.99996. The mean absolute change in hospitals' RSMRs when adding a race indicator is -.01023% with a correlation coefficient between RSMRs for each hospital with and without race added of 0.98496. The mean absolute change in hospitals' RSMRs when adding a low SES AHRQ indicator is 0.00036% with a correlation coefficient between RSMRs for each hospital with and without low SES added of 0.9998.

Overall, we find that among the SDS variables that could be feasibly incorporated into this model the relationship between minority status, dual-eligible status, and low SES (AHRQ indicator status) is in the opposite direction than what has been the expressed concern of stakeholders interested in adding such adjustment to the models. We also find that the impact of any of these indicators is very small to negligible on model performance and hospital profiling. Given the controversial nature of incorporating such variables into a risk-model we do not support doing so in a case that is unlikely to affect hospital profiling. Given these findings and complex pathways could explain any relationship between SDS and mortality, which do not all support risk-adjustment, we did not incorporate SDS variables into the measure.

We do think further investigation or lower mortality among dual eligible and black patients will be valuable and may shed light on proposed future modifications to the measure. For instance, we find that black patients have higher rates of hypertensive heart disease as opposed to other mechanisms for heart failure, which is also consistent with the literature (Echols et al. 2006, Kamath et al. 2008, Thomas et al. 2011). Future reevaluation efforts will explore the relationship between SDS and types of heart failure once ICD-10 data is available.

References:

Echols MR, Felker GM, Thomas KL, et al. Racial differences in the characteristics of patients admitted for acute decompensated heart failure and their relation to outcomes: results from the OPTIME-CHF trial. *Journal of cardiac failure.* 2006;12(9):684-688.

Kamath SA, Drazner MH, Wynne J, Fonarow GC, Yancy CW. Characteristics and outcomes in African American patients with decompensated heart failure. *Archives of internal medicine.* 2008;168(11):1152-1158.

Thomas KL, Hernandez AF, Dai D, et al. Association of race/ethnicity with clinical risk factors, quality of care, and acute outcomes in patients hospitalized with heart failure. *American heart journal.* 2011;161(4):746-754.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Approach to assessing model performance (Dataset 3)
During measure development, we computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the development and validation cohort:

Discrimination Statistics
(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic (also called ROC) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome)
(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects. Therefore, we would hope to see a wide range between the lowest decile and highest decile)

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

We tested the performance of the model in all four datasets described in section 1.7.

Citation:
F.E. Harrell and Y.C.T. Shih, Using full probability models to compute probabilities of actual interest to decision makers, *Int. J. Technol. Assess. Health Care* **17** (2001), pp. 17–26.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
For the development cohort (Dataset 3) the results are summarized below:
1st half of randomly split development sample: c-statistic = 0.71; Predictive ability (lowest decile %, highest decile %) = (3.0, 28.5)
2nd half of randomly split development sample: c-statistic = 0.70; Predictive ability (lowest decile %, highest decile %) = (2.8, 29.0)

For the current measure cohort (Dataset 1) the results are summarized below:
c-statistic = 0.68
Predictive ability (lowest decile %, highest decile %) = (3.7, 27.3)

For the medical record validation cohort (Dataset 2) the results are summarized below:
c-statistic = 0.78
Predictive ability (lowest decile %, highest decile %) = (1.8, 42.4)

c-statistic (linked administrative model validation) = 0.70
Predictive ability (lowest decile %, highest decile %) (linked administrative model validation) = (2.9, 28.4)

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

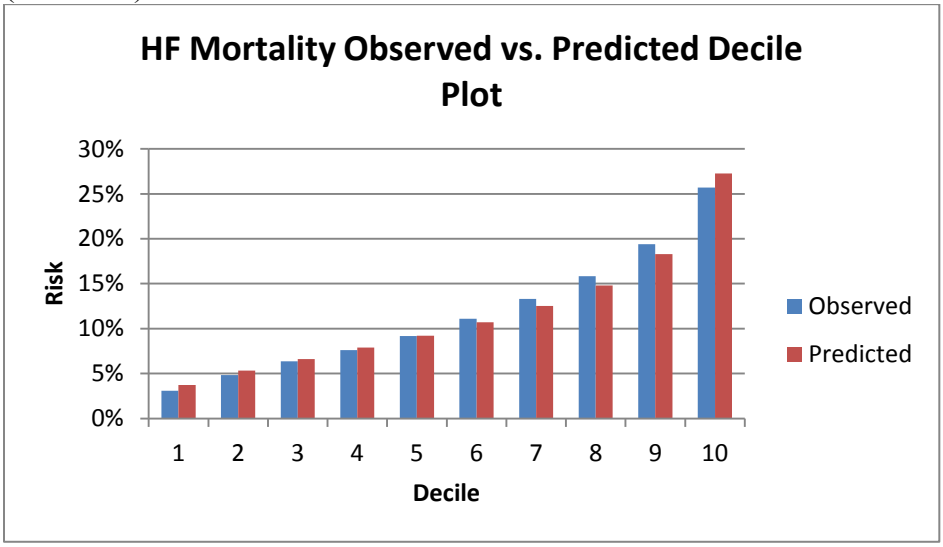
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

For the development cohort (Dataset 3) the results are summarized below:
1st half of split sample: Calibration: (0.0000, 1.0000)
2nd half of split sample: Calibration: (-0.0035, 0.9928)

For the medical record validation cohort (Dataset 2) the results are summarized below:
Calibration: (0.0000, 1.0000)
Calibration (linked administrative model validation): (-0.0045, 1.0021)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

The risk decile plot is a graphical depiction of the deciles calculated to measure predictive ability. Below, we present the risk decile plot showing the distributions for Medicare FFS data from July 2011 to June 2014 (Dataset 1).



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Discrimination Statistics
The c-statistics of 0.68 indicate fair model discrimination (Dataset 1). The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics
Over-fitting (Calibration γ0, γ1)

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

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

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

2b4.11. 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*)

This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSMRs accounting for differences in hospital case-mix.

Approach to assessing model performance:
During measure development, we computed five summary statistics for assessing model performance (Harrell and Shih 2001) for the development and validation cohort:
(1) over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)
(2) predictive ability
(3) area under the receiver operating characteristic (ROC) curve
(4) distribution of residuals
(5) model chi-square (A test of statistical significance usually employed for categorical data to determine whether there is a good fit between the observed data and expected values; i.e., whether the differences between observed and expected values are attributable to true differences in characteristics or instead the result of chance variation).

Application to Medicare FFS Beneficiaries Using Inpatient Data Only for Risk Adjustment (Dataset 4)
As part of testing the model in all-payer data, we also applied the model to CMS data for Medicare FFS 65+ patients in California hospitals using only inpatient data for risk adjustment. Specifically, we created a 2006 measure cohort with complete one-year history data and 30-day follow-up data (N= 24,035).

To help determine whether the measure could be applied to Medicare FFS 65+ patients using only Medicare Part A data, we performed analyses to assess how the model performs when using only admission claims data for risk adjustment, as all-payer hospital discharge databases do not have outpatient claims. To assess the validity of using only admission claims data for risk adjustment, we fit the model separately using the full data and using only admission claims data and (a) compared the odds ratios (ORs) for the various risk factors; (b) conducted a reclassification analysis to compare risk prediction at the patient level; (c) compared model performance in terms of the c-statistic (discrimination); and (d) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and R2) to assess whether the model with only admission claims data is different from the current model in profiling hospital rates.

Adjustment Using CMS data for Medicare FFS 65+ beneficiaries in California hospitals: (a) the magnitude of odds ratios for most risk factors was similar when comparing the model using full data and using only admission claims data; (b) when comparing the model with full data and with only admission claims data, the

reclassification analysis demonstrated good patient-level risk prediction; (c) the c-statistic was similar (0.681 vs. 0.684); and (d) hospital-level risk-standardized rates were highly correlated (r=0.993).

Application to Patients Aged 18 and Older (Dataset 4)
We also applied the model to all-payer data from California. The analytic sample included 60,022 cases aged 18 and older in the 2006 California Patient Discharge Data. When used in all-payer data, only admission claims data are used for risk adjustment, as the hospital discharge databases do not have outpatient claims.

To help determine whether the measure could be applied to an population of patients aged 18+, we examined the interaction terms between age (18-64 vs. 65+) and each of the other risk factors. Specifically, we fit the model in all patients 18+ with and without interaction terms and (a) conducted a reclassification analysis to compare risk prediction at the patient level; (b) compared the c-statistic; and (c) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and R2) to assess whether the model with interactions is different from the current model in profiling hospital rates.

When the model was applied to all patients 18 and over (18+), overall discrimination was good (c-statistic=0.718). In addition, there was good discrimination and predictive ability in both those aged 18-64 and those aged 65+. Moreover, the distribution of Pearson residuals was comparable across the patient subgroups. When comparing the model with and without interaction terms, (a) the reclassification analysis demonstrated good patient-level risk prediction (1.9% to 25.4% vs. 2.0% to 25.1%, respectively, from the bottom decile to the top decile of the prediction values); (b) the c-statistic was nearly identical (0.720 vs. 0.718); and (c) hospital-level risk-standardized rates were highly correlated (r=1.000). Thus, the inclusion of the interactions did not substantively affect either patient-level model performance or hospital-level results.

Therefore, the measure can be applied to all-payer data for patients 18 and older.

References:
Harrell FE, Shih YCT. Using full probability models to compute probabilities of actual interest to decision makers. Int J Technol Assess Health Care. 2001;17:17–26.

Krumholz HM, Normand S-LT, Galusha DH, Mattera JA, Rich AS, Wang YF, Wang Y. et al. Risk-Adjustment Models for AMI and HF: 30-Day Mortality: Report prepared for the Centers for Medicare & Medicaid Services; 2005. Available at: http://www.qualitynet.org/

Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (YNHHSC/CORE) (January 2012). Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following Pneumonia Hospitalization. In *Testing Publicly Report 30-Day Acute Myocardial Infarction, Heart Failure, and Pneumonia Risk-Standardized Mortality and Readmission Measures in California All-Payer Data.*

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

For public reporting of the measure, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national observed mortality rate (is lower or higher than the

rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare website as "better than the U.S. national rate" or "worse than the U.S. national rate." If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or "the difference is uncertain." CMS does not classify performance for hospitals that have fewer than 25 cases in the three-year period.

2b5.2. 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*)

Recent analyses of Medicare FFS data show substantial variation in RSMRs among hospitals. Using data from July 2011-June 2014 (Dataset 1), the median hospital RSMR was 11.7%, with a range of 7.0% to 19.3%. The interquartile range was 10.9%-12.4%.

Out of 4,771 hospitals in the U.S., 145 performed "better than the U.S. national rate," 3,662 performed "no different from the U.S. national rate," and 93 performed "worse than the U.S. national rate." 871 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing. These numbers were reported in the 2015 Condition-Specific Measures Updates and Specifications Report: Hospital-Level 30-Day Risk-Standardized Mortality Measures, prior to exclusion of LVAD, history of LVAD, organ transplantation, and history of organ transplantation.

2b5.3. 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?*)

Despite recent decreases in mortality rates nationally, the mortality rate for HF remains high at 11.7%.

The variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for heart failure that support measurement to reduce the variation.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

N/A

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

N/A

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

N/A

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in

electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Administrative data are routinely collected as part of the billing process.

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)*.
There are no fees associated with the use of this measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

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

Planned	Current Use (for current use provide URL)
	Public Reporting Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html Payment Program

	Hospital Value Based Purchasing (HVBP) Program http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html

4a.1. For each CURRENT use, checked above, provide:

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

Public Reporting

1. Program Name, Sponsor: Hospital Inpatient Quality Reporting (IQR) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was originally mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. Initially, the MMA provided for a 0.4 percentage point reduction in the annual market basket (the measure of inflation in costs of goods and services used by hospitals in treating Medicare patients) update for hospitals that did not successfully report. The Deficit Reduction Act of 2005 increased that reduction to 2.0 percentage points.

In addition to giving hospitals a financial incentive to report the quality of their services, the hospital reporting program provides CMS with data to help consumers make more informed decisions about their health care. Some of the hospital quality of care information gathered through the program is available to consumers on the Hospital Compare website at: www.hospitalcompare.hhs.gov.

Geographic area and number and percentage of accountable entities and patients included: The IQR program includes all IPPS non-federal acute care hospitals and VA hospitals in the United States. The number and percentage of accountable entities included in the program, as well as the number of patients included in the measure, varies by reporting year. For 2015 public reporting, the RSMR will be reported for 4,775 hospitals across the US. The final index cohort includes 991,007 admissions.

2. Program Name, Sponsor: Hospital Value-Based Purchasing (HVBP) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The Hospital Value-Based Purchasing (VBP) Program is a CMS initiative that rewards acute-care hospitals with incentive payments for the quality of care they provide to people with Medicare. It was established by the Affordable Care Act of 2010 (ACA), which added Section 1886(o) to the Social Security Act. The law requires the Secretary of the Department of Health and Human Services (HHS) to establish a value-based purchasing program for inpatient hospitals. To improve quality, the ACA builds on earlier legislation—the 2003 Medicare Prescription Drug, Improvement, and Modernization Act and the 2005 Deficit Reduction Act. These earlier laws established a way for Medicare to pay hospitals for reporting on quality measures, a necessary step in the process of paying for quality rather than quantity.

Geographic area and number and percentage of accountable entities and patients included: More than 3,000 hospitals across the country are eligible to participate in Hospital VBP. The program applies to subsection (d) hospitals located in the 50 states and the District of Columbia and acute-care hospitals in Maryland. Hospital VBP is based on data collected through the Hospital Inpatient Quality Reporting (IQR) Program. More details about the Hospital IQR program are online at https://www.cms.gov/HospitalQualityInits/08_HospitalRHQDAPU.asp.

The following hospitals are excluded from Hospital VBP:
• Hospitals and hospital units excluded from the Inpatient Prospective Payment System, such as psychiatric, rehabilitation, long-term care, children's, and cancer hospitals;
• Hospitals that do not participate in Hospital IQR during the Hospital VBP performance period;
• Hospitals cited by the Secretary of HHS for deficiencies during the performance period that pose an immediate jeopardy to patients' health or safety; and
• Hospitals that do not meet the minimum number of cases, measures, or surveys required by Hospital VBP.

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

N/A, this measure is currently publicly reported

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

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

There has been significant progress in 30-day RSMR for HF. The median 30-day RSMR decreased by 0.7 absolute percentage points from 2011-2012 (median RSMR: 11.7%) to 2013-2014 (median RSMR: 11.0%). The median hospital RSMR from 2011-2014 was 11.7% (IQR 11.0% - 12.5%). In addition, hospitals with a high proportion of Medicaid and African American patients achieve a similar range of performance as compared with hospitals with a low proportion of these patients, indicating that both groups of hospitals can perform well on the measure.

4b.2. **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.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We did not identify any unintended consequences during measure development or model testing. However, we are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative unintended consequences for patients.

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)**
0230 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older
0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
0358 : Heart Failure Mortality Rate (IQI 16)

0468 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization
0505 : Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.
0506 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
1551 : Hospital-level 30-day, all-cause risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
1891 : Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization
1893 : Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization

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

5a. Harmonization

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 completely harmonized?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
We did not include in our list of related measures any non-outcome (e.g., process) measures with the same target population as our measure. Our measure cohort was heavily vetted by clinical experts, a technical expert panel, and a public comment period. Additionally, the measure, with the specified cohort, has been publicly reported since 2008. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

5b. Competing Measures

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

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

Appendix

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

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
Co.2 **Point of Contact:** Lein, Han, Lein.han@cms.hhs.gov, 410-786-0205-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)
Co.4 **Point of Contact:** Lisa, Suter, Lisa.suter@yale.edu, 203-737-3400-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org.

Our measure development team consisted of the following members:

Kanchana R. Bhat, M.P.H., Project Coordinator
Elizabeth E. Drye, M.D., S.M., Project Director
Harlan M. Krumholz, M.D., S.M., Principal Investigator
Sharon-Lise T. Normand, Ph.D., Co-Investigator*
Geoffrey C. Schreiner, B.S., Research Assistant
Yongfei Wang, M.S., Senior Statistical Analyst
Yun Wang, Ph.D., Senior Biostatistician

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2008
Ad.3 **Month and Year of most recent revision:** 04, 2015
Ad.4 **What is your frequency for review/update of this measure?** This measure is updated annually.
Ad.5 **When is the next scheduled review/update for this measure?** 04, 2016

Ad.6 **Copyright statement:** N/A
Ad.7 **Disclaimers:** N/A

Ad.8 **Additional Information/Comments:** N/A

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 #: 0230
Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older
Measure Steward: Centers for Medicare & Medicaid Services (CMS)
Brief Description of Measure: The measure estimates a hospital-level 30-day risk-standardized mortality rate (RSMR). Mortality is defined as death for any cause within 30 days after the date of admission for the index admission, for patients 18 and older discharged from the hospital with a principal diagnosis of acute myocardial infarction (AMI). CMS annually reports the measure for patients who are 65 years or older and are either Medicare fee-for-service (FFS) beneficiaries and hospitalized in non-federal hospitals or are hospitalized in Veterans Health Administration (VA) facilities.
Developer Rationale: The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized mortality rates following hospitalization for AMI. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This measure was developed to identify institutions whose performance is better or worse than what would be expected based on their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

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

Numerator Statement: The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days of the index admission date for patients 18 and older discharged from the hospital with a principal diagnosis of AMI.
Denominator Statement: This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 18 years or older. The cohort includes admissions for patients discharged from the hospital with a principal discharge diagnosis of AMI and with a complete claims history for the 12 months prior to admission. Currently, the measure is publicly reported by CMS for those patients 65 years and older who are either Medicare FFS beneficiaries admitted to non-federal hospitals or patients admitted to VA hospitals. Additional details are provided in S.9 Denominator Details.
Denominator Exclusions: The mortality measures exclude index admissions for patients:
1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility.
2. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission; or
4. Discharged against medical advice (AMA).

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

For Medicare FFS patients, the measure additionally excludes admissions for patients without at least 30 days post-discharge

enrollment in FFS Medicare (because the 30-day mortality outcome cannot be assessed in this group).

Measure Type: Outcome
Data Source: Administrative claims, Other, Paper Medical Records
Level of Analysis: Facility
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: 5/9/07 **Most Recent Endorsement Date:** 1/18/12

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

Public and Member Comments on original measure

- All-cause mortality rate does not correlate well with AMI mortality.

Committee response:

- All patient care is inter-related. All-cause mortality reflects the reality of caring for patients. It is not possible to separate "cardiovascular causes" independent of other conditions affecting a patient

Public and Member Comments on revised measure:

- Clarify the data sources that were used in the all payer data testing.

Developer Response:

- The data source used to complete the all payer testing was the state of California's Patient Discharge Database (PDD) which contains records for all discharges from all non-Federal hospitals located in California. California is a diverse state, and, with more than 37 million residents, California represents 12% of the US population. In 2006, there were approximately 3 million adult discharges from more than 450 hospitals. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations and to evaluate rates of both readmission and mortality. Specifically, patients from this database are linked to the California Death Statistical Master File (DSMF) using social security number in order to validate and record deaths.

Steering Committee: Agreed that the developer answered the comment.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing a rationale that supports the relationship of the outcome to at least one healthcare structure, process, intervention, or service.

- This measure calculates hospitals' 30-day risk-standardized mortality rate (RSMR) for patients who have been hospitalized with acute myocardial infarction (AMI).
- As a rationale for measuring this health outcome, the developer suggests that hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of complications, use of appropriate medications, timely percutaneous coronary interventions, discharge planning, management of care transitions, medication reconciliation, and patient education.
- The developer states that recent qualitative research funded by AHRQ, Commonwealth Fund, and United Healthcare identified common system-level approaches to care and, specifically, the tailored use of protocols, in those hospitals that have low RSMRs compared with hospitals with high RSMRs.
- The developer has included a number of references to support this rationale.

Question for the Committee:

- *Does the Committee agree that hospitals have the ability to influence 30-day mortality rates among AMI patients?*

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

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

- The developer notes that AMI mortality is a priority area for outcomes measure development as it is a costly and common condition.
- The developer provides performance data from four measurement periods, covering a total of 497,550 admissions.
- The data show that during the measurement period of 07/2011–06/2014, AMI mortality rates ranged from a minimum of 9.9% to a maximum of 20.6%, with the 10th percentile at 13.0%, the 50th percentile at 14.2%, and the 90th percentile at 15.4%.
- To help in assessment of potential disparities, the developers also provide performance scores (using 2011-2014 data) for hospitals serving a low proportion of Medicaid patients vs. those serving a high proportion of Medicaid patients, and performance scores for hospitals serving a low proportion of African-American patients vs. those serving a high proportion of African-American patients.
- By proportion of **Medicaid Patients**:

  // **Low proportion (=8.4%) Medicaid patients** // **High proportion (=30.5%) Medicaid patients**
  Number of Measured Entities (Hospitals)// 242 // 243
  Number of Patients// 49,022 in low-proportion hospitals // 37,060 in high-proportion hospitals
  Maximum// 18.9 // 18.1
  90th percentile// 15.3 // 15.7
  75th percentile// 14.6 // 15.1
  Median (50th percentile)// 13.7 // 14.3
  25th percentile// 13.1 // 13.5
  10th percentile// 12.5 // 12.9
  Minimum // 10.0 // 11.0

- By proportion of **African-American patients**:

  // **Low Proportion (=0%) Af-Am patients** // **High proportion (=23.7%) Af-Am patients**
  Number of Measured Entities (Hospitals)// 244 // 245
  Number of Patients// 28,674 patients in low-proportion hospitals//38,275 in high-proportion hospitals
  Maximum// 17.5 // 17.9
  90th percentile// 16.0 // 15.8
  75th percentile// 15.2 // 15.1
  Median (50%)// 14.3 // 14.2
  25th percentile// 13.5 // 13.4
  10th percentile// 12.7 // 12.7
  Minimum// 11.5 // 10.4

- The developers do not provide interpretation or analysis of these data; in both cases there are differences in performance between the groups compared, but these differences do not appear to be substantial.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - Sufficient evidence was presented - and has previously been presented - to justify monitoring this risk-adjusted outcome

1a. ***Committee's Comments on Evidence to Support Measure Focus:***
   - Strong evidence.

1b. ***Committee's Comments on Performance Gap:***
   - Large gap in performance
   - The variability in this outcome is substantial across hospitals

1c. ***Committee's Comments on Composite performance Measure:***
   - Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- This measure calculates 30-day all-cause mortality for patients hospitalized with acute myocardial infarction (AMI).
- The measure produces a risk-standardized mortality rate (RSMR), which is calculated as the ratio of the number of "predicted" to the number of "expected" deaths, multiplied by the national unadjusted mortality rate.
- The denominator includes patients aged 18 years and older discharged from the hospital with a principal discharge diagnosis of AMI and with a complete claims history for the 12 months prior to admission. The measure can also be calculated for patients aged 65 and older only.
- The numerator includes patients who died of any cause within 30 days of the date of admission of the index AMI hospitalization.
- The denominator population is defined using ICD-9 and ICD-10 codes; a list of applicable codes is included in the submission.
- The numerator population is defined using vital status data, which may be derived from the Medicare Enrollment Database (EDB), State-based data systems, the Social Security Administration's Death Master File (DMF), or the Centers for Disease Control and Prevention's National Death Index (NDI).
- The data sources for this measure may include Medicare Part A and B claims, Veterans Health Administration claims, the Medicare Enrollment Database (EDB), and all-payer data sources such as the California Patient Discharge Database.
- The measure's time window can be specified from one to three years.
- The measure is risk-adjusted using a statistical risk model (see details below).

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- The developer has assessed reliability at both the data element and the performance score levels.
- Datasets used for testing included Medicare Parts A and B claims, Veterans' Health Administration claims, as well as the Medicare Enrollment Database (EDB). Additionally, census data were used to assess socio-demographic factors.
    - **Data element reliability**:
        - With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
        - In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time. [*Note: NQF does not typically consider temporal consistency to be a valid method of demonstrating reliability of data elements*.]
        - Summarizing the results of this analysis, the developer notes that the frequency of some model variables increased between 2011 and 2014, which may reflect increased co-morbidity rates, but may also be due to increased hospital coding of comorbidities.
        - The developer notes that as part of the 2012 update to this measure, the number of diagnosis codes and procedure codes were increased to align with the Version 5010 format changes required by the Department of Health and Human Services (DHHS).
    - **Performance score reliability**:
        - The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
        - In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. [*Note: NQF considers this to be an appropriate method of assessing reliability*.]
        - A total of 497,550 admissions over a 3-year period were examined, with 247,641 in one sample and 249,909 in the other randomly-selected sample. Two risk-standardized mortality rates (RSMR) were calculated for each hospital: one from each of the two separate samples.
        - The agreement between the two RSMRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was **0.41**; the developer states that according to the conventional interpretation, this is considered a "moderate" level of agreement.
        - The developer notes that this analysis was limited to hospitals with 12 or more cases in each split sample, and that splitting the total population into two samples resulted in a sample equivalent of only 1.5 years of data, whereas the measure is reported with the full three years of data. [*Note: It is unclear whether the measure itself is limited to hospitals with 12 or more cases; if it is not, then testing was not conducted with the measure as specified*.]
        - The developer expects that the correlation coefficient would be higher using a full three-year sample since it would include more patients.
- The developer's overall interpretation of reliability testing results is that the stability of the risk factor frequencies and odds ratios over time suggests that the underlying data elements are reliable, and that the ICC score from performance score analysis demonstrates moderate agreement across samples using a conservative approach to assessment.

Questions for the Committee:

- *Do the testing results presented by the developer demonstrate an adequate level of reliability?*
- *In addition to the consistency of measurement results, assessments of performance score reliability often examine the ability of the measure to differentiate between measured entities. Do the reliability testing results reported by the developer demonstrate that meaningful differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- This measure estimates 30-day all-cause mortality for patients hospitalized with acute myocardial infarction (AMI) using a risk standardized mortality ratio (RSMR), which is calculated as the ratio of the number of "predicted" to the number of "expected" deaths, multiplied by the national unadjusted mortality rate.
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of complications, provision of evidenced-based care, discharge planning, management of care transitions, medication reconciliation, and patient education.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- The developer conducted empirical validity testing of the measure score.
- To assess validity, the developer compared scores from the administrative claims-based measure (i.e., the measure as specified) to scores derived from medical record review in the same patient cohort.
- This assessment was conducted on data from 1994 and 1995, comprising 178,188 AMI hospitalizations; the unadjusted 30-day mortality rate in this population was 19.0%.
- Hospital-level risk-standardized mortality rates (RSMRs) were estimated using the claims-based model and the medical record-based model; the linear relationship between these two sets of estimates was then examined, using regression techniques and weighting by the total number of cases in each hospital.
- The correlation between the claims-based RSMRs and the record-based RSMRs was estimated at 0.91, which the developer characterizes as a "strong" level of correlation.

Questions for the Committee:

- *Do the method and results of testing demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients in the following categories are excluded from the measure:
- Discharged alive on the day of admission or the following day who were not transferred to another acute care facility.
- With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
- Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission; or
- Discharged against medical advice (AMA).
- In addition, patients without at least 30 days post-discharge enrollment in FFS Medicare are excluded from the measure (because the 30-day mortality outcome cannot be assessed in this group).
- The developer notes that all exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure
- To determine the impact of exclusions, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- The number and percentage of patients excluded for each criterion are as follows:

1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility: **34,657 (6.17%)**
2. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data: **26 (<0.01%)**
3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission: **4,770 (0.85%)**
4. Discharged against medical advice (AMA): **2,630 (0.47%)**

- The developer also provides the distribution across hospitals for each exclusion criterion.
- The developer notes that the first exclusion criterion, which accounts for the majority of all exclusions, is meant to ensure a clinically coherent cohort, preventing the inclusion of patients who likely did not suffer a clinically-significant AMI.
- The developer comments that for most hospitals, this results in very few patients being excluded, and that for those hospitals with greater proportions of excluded patients, the measure is likely excluding less severe patients that may not be considered as AMI at other hospitals. The developer notes that this exclusion was guided by the input of clinical experts at the time of measure development.
- The developer states that exclusion criteria 2 and 3 are necessary for valid calculation of the measure, and that exclusion 4 is needed for acceptability of the measure to hospitals who do not have the opportunity to adequately prepare such patients for discharge, noting that the aforementioned exclusions are unlikely to affect the measure score because they exclude a very small percentage of patients.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital level 30-day risk-standardized mortality rate (RSMR).
- The developer suggests that this approach to modeling appropriately accounts for the structure of the data (patients clustered within hospitals), the underlying risk due to patients' comorbidities, and the sample size at a given hospital when estimating hospital mortality rates
- The developer notes that this risk-adjustment approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes both within and between hospitals.
- Variables considered for inclusion in the model were patient-level risk-adjustors that are expected to be predictive of mortality based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity.
- For each patient, covariates were obtained from claims records extending 12 months prior to and including the index admission. The covariates are defined using condition categories (CCs), which are clinically-meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes.
- The measure does not adjust for CCs that were possible adverse events of care and that were only recorded in the index admission.
- The final set of 27 risk-adjustment variables is included in the testing attachment; the odds ratio associated with each variable is also provided.

- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.

- **Conceptual analysis of the need for SDS adjustment:**
  - The developers note that there is a large body of literature linking various SDS factors to worse health status and higher mortality over a lifetime, with income, education, and occupational level being the

most commonly examined variables, though literature directly related to 30-day mortality after admission for cardiovascular disease is much more limited.
    - One potential pathway for SDS factors to affect 30-day mortality (independent of the quality of care) is patients' health status at the time of admission.
        - SDS factors can influence admission health status both due to the impact of multiple related stressors over a lifetime contributing to overall worse health, as well as through poorer access to care and potentially delayed presentation.
        - The developers note that this pathway should be largely accounted for by their clinical risk-adjustment model.
    - Another potential pathway for SDS factors to affect 30-day mortality is for lower-income patients to elect not to follow prescribed care (for example, refill a prescription or keep a follow-up visit with a primary care provider) because limited resources create competing priorities for the patient.
    - The developers also argue that there are a number of pathways for SDS to affect 30-day mortality that are *not* independent of the quality of care. These may include:
        - Contextual effects, such as patients of low SDS presenting at lower-quality institutions for care;
        - Patients of low SDS receiving differentiated care as compared to counterparts of higher SDS—which may be appropriate or inappropriate in different instances.

- **Empirical analysis of SDS factors:**
    - The developers state that they found race (black vs non-black) and dual-eligible status—i.e., enrolled in both Medicare and Medicaid (obtained from CMS claims enrollment data)—to be the only two patient-level SDS variables available for direct examination.
    - Also considered were a number of neighborhood-level variables (represented in a validated AHRQ composite index of SDS variables found in census data, including income and education) that could serve as a proxy for patient-level SDS.
        - These variables are linked to patients by zip code; however, the data are only linked at a 5-digit zip code level—nine-digit zip code data, which may provide a more granular view of patient sociodemographic status, were not available.
        - Patients were identified as low SDS if they lived in a neighborhood in the lowest quartile of the AHRQ SDS index.
    - The developer's method was to first evaluate variation in the prevalence of low-SDS patients among providers; they then assessed the relationship (univariate) between the SDS variables and 30-day HF mortality, and examined the incremental effect of SDS in a multivariable model, evaluating the extent to which the addition of any one of these variables improved model performance or changed hospital results.
    - The developers' analysis found that the prevalence of SDS factors in the HF cohort does vary across measured entities.
    - With regard to the empirical association of each SDS variable with the outcome (univariate), the analysis found that patient-level observed AMI unadjusted mortality rate for dual-eligible patients was somewhat higher, at 16.1% compared with 14.0% for all other patients; the mortality rate for black patients was lower at 12.6% compared with 14.4% for patients of all other races, and the mortality rate for patients in the lowest SES quartile by AHRQ Index was slightly higher at 14.4% compared with 13.9% for patients in the highest SES quartile.
        - With regard to the strength and significance of the SDS variables in the context of a multivariable model, the developers' analysis found that:
            - For dual eligibility and the AHRQ SES indicator, the variable is associated with higher risk of modest strength. Odds ratios are on the order of 1.12 for dual eligibility and 1.09 for AHRQ SES.
            - The developer notes that this is similar to the odds ratio for comorbidities such as COPD and substantially lower than the risk associated with comorbidities such as metastatic cancer.
            - For race, black race is associated with a lower risk of mortality, with an odds ratio of 0.81.
        - However, the developer states that in all cases, the c-statistic (i.e., predictive value) for the AMI patient-level multivariate model with the SDS variable in the model is essentially unchanged from that without.
        - To further understand the relative importance of these risk-factors in the measure, the developers compared hospital performance with and without the addition of each SDS variable.

    - Their analysis found that the addition of any of these variables into the model had little to no effect on hospital performance.
- Regarding their overall findings, the developers state that among the SDS variables that could be feasibly incorporated into this model,
    - The relationship with mortality is small;
    - The relative effect of black race is stronger than the other two SDS variables but is in the opposite direction than what has been the expressed concern of stakeholders interested in adding such adjustment to the models;
    - The impact of adding any of these indicators is very small to negligible on model performance and hospital profiling.
- The developers state that given these findings and complex pathways that could explain any relationship between SDS and mortality, which do not all support risk-adjustment, **they did not incorporate SDS variables into the measure**.

- **Risk Model Diagnostics:**
    - To assess the overall performance of their risk-adjustment model, the developers computed several summary statistics, including:
        - *Area under the receiver operating characteristic (ROC) curve* (also known as a c-statistic, which measures the probability that the model's prediction of the outcome is better than chance)
        - *Predictive ability* (the model's ability to distinguish high-risk subjects from low-risk subjects)
        - *Over-fitting indices (model calibration)* (to ensure that the model is not only describing the relationship between predictive variables and outcome in the development dataset but also providing valid predictions in new patients)
    - For the current measure cohort, the findings from this analysis are as follows:
        - C-statistic: **0.72**
            - A c-statistic of 0.72 means that for 72% of all possible pairs of patients—one who died and one who lived—the model correctly assigned a higher probability to those who died. Generally, a c-statistic of at least 0.70 is considered acceptable.
            - The developers interpret this as 'good' model discrimination
        - Predictive ability (lowest decile %, highest decile %): (**2.8%, 33.3%**)
            - The developers state that this indicates a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.
        - Overfitting indices (model calibration) [presented as (γ0, γ1)]:
            - The developer states that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.
                - 1st half of split sample: Calibration: (**0.0000, 1.0000**)
                - 2nd half of split sample: Calibration: (**-0.0035, 0.9928**)
    - The developer's overall interpretation of the results of their analysis is that the findings demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

    - The developer also conducted additional analyses to determine whether the measure could be applied to Medicare FFS 65+ patients using only Medicare Part A data and whether it could be applied to a population of patients aged 18+ using all-payer data.
    - The developers report that their results indicate their model had good discrimination and predictive ability in both groups.

Questions for the Committee:

- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*

- *Does the Standing Committee agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*
- *Do you agree with the developer's decision, based on their empirical analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference:

- For public reporting of this measure, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate.
- If the RSMR's interval estimate does not include the national observed mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare website as "better than the U.S. national rate" or "worse than the U.S. national rate."
- If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or "the difference is uncertain."
- The developer reports that for the performance period of July 2011-June 2014, the mean hospital RSMR was 14.3%, with a range of 9.9% to 20.6%. The interquartile range was 13.8%-14.8%.
- Of 4,490 hospitals in the study cohort, 41 performed "better than the U.S. national rate," 2,474 performed "no different from the U.S. national rate," 21 performed "worse than the U.S. national rate," and 1,954 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.
- The developer's interpretation of this data is that the variation in rates and number of performance outliers suggests there remain differences in the quality of care received across hospitals for AMI that support measurement to reduce the variation.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- While the developer did not decide to include SDS variables in their final model, they did compare measure results with and without SDS adjustment.

2b7. Missing Data

- N/A

Committee pre-evaluation comments

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

2a1. & 2b1.: ***Committee's Comments on Reliability-Specifications:***

- Clearly defined administrative claims data
- "The developers do not mention a potential lag in the reporting of numerator data. Also - the verbiage below could cause concern in terms of definitions:
  - Summarizing the results of this analysis, the developer notes that the frequency of some model variables increased between 2011 and 2014, which may reflect increased co-morbidity rates, but may also be due to increased hospital coding of comorbidities.
  - The developer notes that as part of the 2012 update to this measure, the number of diagnosis codes and procedure codes were increased to align with the Version 5010 format changes required by the Department of Health and Human Services (DHHS).
  - [Note: It is unclear whether the measure itself is limited to hospitals with 12 or more cases; if it is not, then testing was not conducted with the measure as specified.]

2a2.: ***Committee's Comments on Reliability-Testing:***

- Intra-class correlation coefficient of 0.41 between RSMRs for same hospital in 1.5 year periods back to back (split half method). Just barely acceptable
- The measure appears to be reliably reproducible,

2b1.: ***Committee's Comments on Validity-Specifications*** :

- Clearly defined administrative claims data
- None evident

2b2.: ***Committee's Comments on Validity-Testing:***

- Intraclass coefficient of 0.91 compared to medical record review in large sample
- Reliability testing would have been more meaningful if the developers had demonstrated (as they did for SDS factors) differences in RSMRs between the clinical and the claims model

2b3-7.: ***Committee's Comments on Threats to Validity:***

- They opted not to use SDS, but at least did extensive testing to show that the impact of candidate variables on the model was very little
- "2b3-Some sites appear to have 100% exclusions - that would be worrisome
    - 2b4 - the model was generated on a California population, which may not be representative - it is curious that a model on the entire population was not developed
    - 2b5 - meaningful differences between top and bottom percentiles are apparent
    - 2b6 - the results presented are supportive of not including the SDS factors
    - 2b7 - developers maintain that missing values are rare but do not demonstrate the extent of missing values

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- This measure is based on administrative claims data (e.g., DRG, ICD-9/10), which the developers note are routinely generated and collected as part of hospitals' billing processes.
- The developer indicates that all data elements are in defined fields in electronic claims.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility***

- Claims data, very feasible
- The measure should be readily feasible

Criterion 4: Usability and Use

4. Usability and 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.

- Measure results are publicly reported through CMS's Hospital Inpatient Quality Reporting (IQR) Program.
- In addition, measure results are incorporated into the calculation of hospital payment rates through CMS's Hospital Value Based Purchasing (HVBP) Program.
- The developer reports that there has been significant progress in 30-day mortality rates for AMI, noting that the median 30-day RSMR decreased by 1.4% from 2011-2012 (median RSMR: 14.7%) to 2013-2014 (median RSMR:

13.3%).

- The developers did not identify any unintended consequences during measure development or model testing, but note that they are committed to monitoring this measure's use and assessing potential unintended consequences over time.

***Questions for the Committee**:*

- *Are this measure's performance results being used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- Publicly reported and should be
- The measure is currently being used by several programs.

Criterion 5: Related and Competing Measures

- List any related or competing measures based on harmonization protocol.
- Summarize any harmonization efforts, i.e., responses from the developers regarding harmonization.
- Briefly summarize next steps according to protocol

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older

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

Date of Submission: Click here to enter a date

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.

- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

1a.1.This is a measure of:

Outcome

☒ Health outcome: cost/resource use.

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

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

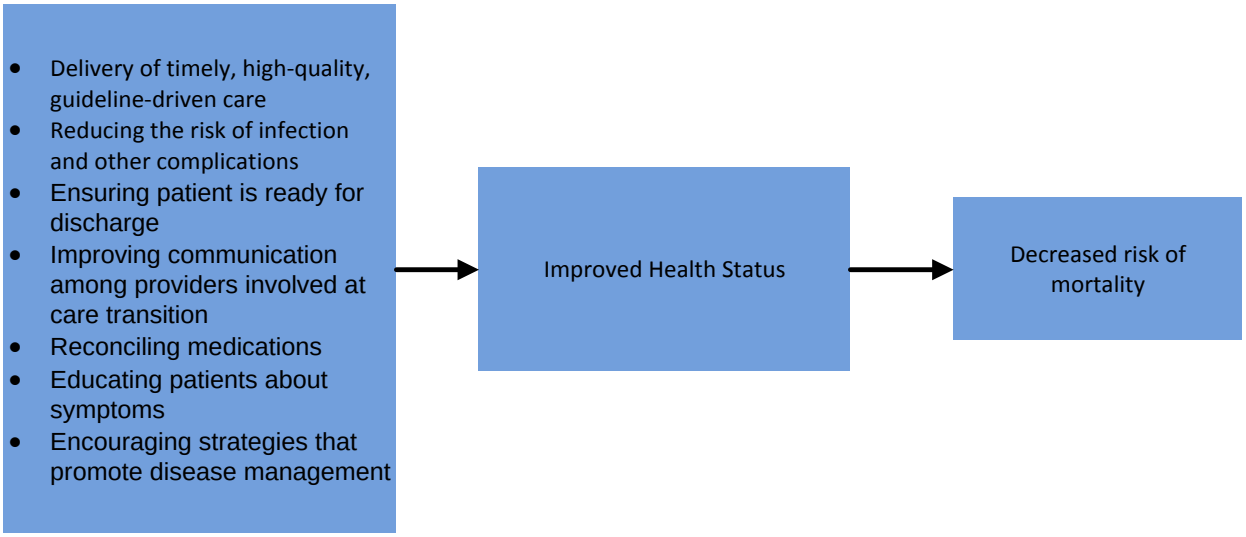
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



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

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Acute myocardial infarction (AMI) is one of the most common principal hospital discharge diagnoses among older adults and is associated with high mortality. Approximately 635,000 Americans will have AMI and approximately 280,000 will have a recurrent attack (Go et al. 2013). It is estimated that an additional 150,000 AMIs occur each year, estimating a total of 1,065,000 AMI-related events a year (Go et al. 2013). The estimated acute myocardial infarction (AMI) incidence in 2011 was 610,000 new attacks and 325,000 recurrent attacks (National Quality Measures). Additionally, AMI was the tenth most common principal discharge diagnosis among patients with Medicare in 2012 (AHRQ 2010).

The high prevalence and considerable morbidity and mortality associated with AMI create an economic burden on the healthcare system (American Heart Association, 2010). AMI accounts for a large fraction of hospitalization costs, and was the sixth most expensive condition billed to Medicare in 2011 (Torio et al 2013). It is estimated that in 2009 the combination of direct and indirect health care costs of coronary Heart disease reached over $195.2 billion (Go et al. 2013).

Hospital interventions, such as use of appropriate medications, timely percutaneous coronary interventions and prevention of complications, are known to decrease the risk of death within 30 days of hospital admission (Rathore et al. 2009; Antman et al. 2008; Jha et al. 2007).Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals.

Over the last 10 years, nationally, risk-standardized mortality rates have decreased for AMI (Krumholz et al. 2009). Yet, continued variation in performance suggests continued opportunities for improvements. In addition, recent qualitative research funded by AHRQ, Commonwealth Fund, and United Healthcare identified common system-level approaches to care and, specifically, the tailored use of protocols in those hospitals that have low RSMRs compared with hospitals with high RSMRs (Curry et al. 2011).

References:

Agency for Healthcare Research and Quality (AHRQ). *Healthcare Cost and Utilization Project (HCUP)* http://hcupnet.ahrq.gov/.

American Heart Association. Heart Disease and Stroke Statistics – 2010 Update. Dallas, Texas: American Heart Association; 2010. c2010, American Heart Association.

Andrews RM, Elixhauser A. The national hospital bill: growth trends and 2005 update on the most expensive conditions by payer. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ); 2007 Dec. (HCUP statistical brief; no. 42).

Antman EM, Hand M, Armstrong PW, et al. 2007 focused update of the ACC/AHA 2004 guidelines for the management of patients with ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2008 Jan 15;51(2):210-247.

Curry LA, Spatz E, Cherlin E, Thompson JW, Berg D, Ting HH, Decker C, Krumholz HM, Bradley EH . What Distinguishes Top-Performing Hospitals in Acute Myocardial Infarction Mortality Rates? A Qualitative Study. Ann Intern Med. 2011;154:384-390.

National Quality Measures, C. "Acute myocardial infarction (AMI): mortality rate." Retrieved 4/7/2015, from http://www.qualitymeasures.ahrq.gov/content.aspx?id=38492&search=Mortality+rate.

Go, A. S., et al. (2013). "Heart disease and stroke statistics--2013 update: a report from the American Heart Association." Circulation 127(1): e6-e245.

Krumholz HM, Wang Y, Chen J, Drye EE, Spertus JA, Ross JS, Curtis JP, Nallamothu BK, Lichtman JH, Havranek EP, Masoudi FA, Radford MJ, Han LF, Rapp MT, Straube BM, Normand SL. Reduction in acute myocardial infarction mortality in the United States: risk-standardized mortality rates from 1995-2006. JAMA. 2009 Aug 19;302(7):767-73.

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

Rathore SS, Curtis JP, Chen J, Wang Y, Nallamothu BK, Epstein AJ, Krumholz HM; National Cardiovascular Data Registry. Association of door-to-balloon time and mortality in patients admitted to hospital with ST elevation myocardial

infarction: national cohort study. BMJ. 2009 May 19; 338:b1807 .

Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics— 2013 update: a report from the American Heart Association. Circulation. 2013; 127:e6-e245.

Torio CM, Andrews RM.: National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011. HCUP Statistical Brief #160. 2013; http://www.hcup-us.ahrq.gov/reports/statbriefs/sb160.jsp

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
AMI_mortality_NQF_Evidence_Attachment_06-29-15.docx

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

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

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Distribution of Hospital AMI RSMRs over Different Time Periods:

Characteristic	July 2011- June 2012	July 2012- June 2013	July 2013-June 2014	July 2011-June 2014
Number of Hospitals	4, 166	4,102	3,997	4,490
Number of Admissions	167,291	169,885	160,374	497,550
Mean Number of Admissions	40.2	41.4	40.1	110.8
Range (min. – max.)	11.2-19.4	11.0-19.4	10.9-16.3	9.9-20.6%
Minimum	11.2	11.0	10.9	9.9
10th percentile	13.9	13.7	12.8	13.0
20th percentile	14.3	14.1	13.1	13.6
30th percentile	14.5	14.3	13.2	13.9
40th percentile	14.7	14.4	13.3	14.1
50th percentile	14.7	14.5	13.3	14.2
60th percentile	14.9	14.7	13.4	14.4
70th percentile	15.0	14.9	13.5	14.6
80th percentile	15.2	15.0	13.6	14.9
90th percentile	15.6	15.4	13.9	15.4
Maximum	19.4	19.4	16.2	20.6

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

N/A

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

Distribution of AMI RSMRs by Proportion of Medicaid Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic	Hospitals with a low proportion (=8.4%) Medicaid patients	Hospitals with a high proportion (=30.5%) Medicaid patients
Number of Measures Entities (Hospitals)	242	243
Number of Patients	49,022 in low-proportion hospitals	37,060 in high-proportion hospitals
Maximum	18.9	18.1
90th percentile	15.3	15.7
75th percentile	14.6	15.1
Median (50th percentile)	13.7	14.3
25th percentile	13.1	13.5
10th percentile	12.5	12.9
Minimum	10.0	11.0

Distribution of RSMRs by Proportion of African-American Patients:
Dates of Data: July 2011 through June 2014
Data Source: Medicare FFS claims

Characteristic	Hospitals with a low Proportion (=0%) African-American patients	Hospitals with a high proportion (=23.7%) African-American patients
Number of Measures Entities (Hospitals)	244	245
Number of Patients	28,674 patients in low-proportion hospitals	38,275 in high-proportion hospitals
Maximum	17.5	17.9

90th percentile// 16.0 // 15.8
75th percentile// 15.2 // 15.1
Median (50%)// 14.3 // 14.2
25th percentile// 13.5 // 13.4
10th percentile// 12.7 // 12.7
Minimum// 11.5 // 10.4

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Acute myocardial infarction (AMI) is one of the most common principal hospital discharge diagnoses among older adults and is associated with high mortality. The high prevalence and considerable morbidity and mortality associated with AMI create an economic burden on the healthcare system (American Heart Association, 2010). In 2005, AMI was the fourth most expensive condition treated in US hospitals, accounting for nearly 4% of the national hospital bill. It was also the fourth most expensive condition billed to Medicare that year, accounting for 4.5% of Medicare´s hospital bill (Andrews and Elixhauser, 2007).

Approximately 635,000 Americans will have AMI and approximately 280,000 will have a recurrent attack. It is estimated that an additional 150,000 MIs occur each year, creating an estimated total of 1,065,000 AMI related events a year. It is estimated that the combination of direct and indirect health care costs of coronary heart disease reached over $195.2 billion (2009).

Many current hospital interventions are known to decrease the risk of death within 30 days of hospital admission (Jha et al. 2007; Rathore et al. 2009). Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
American Heart Association. Heart Disease and Stroke Statistics – 2010 Update. Dallas, Texas: American Heart Association; 2010. c2010, American Heart Association.

Andrews RM, Elixhauser A. The national hospital bill: growth trends and 2005 update on the most expensive conditions by payer. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ); 2007 Dec. (HCUP statistical brief; no. 42).

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

Rathore SS, Curtis JP, Chen J, Wang Y, Nallamothu BK, Epstein AJ, Krumholz HM; National Cardiovascular Data Registry. Association of door-to-balloon time and mortality in patients admitted to hospital with ST elevation myocardial infarction: national cohort study. BMJ. 2009 May 19; 338:b1807.

Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics— 2013 update: a report from the American Heart Association. Circulation. 2013; 127:e6-e245.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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 : Acute Myocardial Infarction

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830 &
http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** AMI_Mortality_NQF_Data_Dictionary_06-22-15_FINAL.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Annual Updates
1. Updated CC map.
a. Rationale: The ICD-9-CM CC map is updated annually to capture all relevant comorbidities coded in patient administrative claims data.

No other updates made after 2013 Measure Updates except for use of new years of data for public 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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days of the index admission date for patients 18 and older discharged from the hospital with a principal diagnosis of AMI.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back*

to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)
Numerator time window: We define the time period for death from any cause within 30 days from the date of admission of the index AMI hospitalization.

Denominator time window: This measure was developed with 12 months of data. The time window can be specified from one to three years. Currently, the measure is publicly reported with three years of index hospitalizations.

S.6. **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, 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.
The measure counts deaths for any cause within 30 days of the date of admission of the index AMI hospitalization.

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

Identifying deaths in the all-payer measure
For the purposes of development, deaths were identified using the California vital statistics data file. Nationally, post-discharge deaths can be identified using an external source of vital status, such as the Social Security Administration's Death Master File (DMF) or the Centers for Disease Control and Prevention's National Death Index (NDI).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 18 years or older. The cohort includes admissions for patients discharged from the hospital with a principal discharge diagnosis of AMI and with a complete claims history for the 12 months prior to admission. Currently, the measure is publicly reported by CMS for those patients 65 years and older who are either Medicare FFS beneficiaries admitted to non-federal hospitals or patients admitted to VA hospitals. Additional details are provided in S.9 Denominator Details.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
To be included in the measure cohort used in public reporting, patients must meet the following additional inclusion criteria:
1. Having a principal discharge diagnosis of AMI;
2. Enrolled in Medicare FFS;
3. Aged 65 or over;
4. Not transferred from another acute care facility; and
5. Enrolled in Part A and Part B Medicare for the 12 months prior to the date of index admission, and enrolled in Part A during the index admission.

VA beneficiaries/hospitalizations are also included in the AMI mortality measure. Enrollment in Medicare FFS is not required for these patients.

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the cohort for each measure are:
410.00 AMI (anterolateral wall) – episode of care unspecified
410.01 AMI (anterolateral wall) – initial episode of care
410.10 AMI (other anterior wall) – episode of care unspecified
410.11 AMI (other anterior wall) – initial episode of care
410.20 AMI (inferolateral wall) – episode of care unspecified
410.21 AMI (inferolateral wall) – initial episode of care
410.30 AMI (inferoposterior wall) – episode of care unspecified
410.31 AMI (inferoposterior wall) – initial episode of care

410.40 AMI (other inferior wall) – episode of care unspecified
410.41 AMI (other inferior wall) – initial episode of care
410.50 AMI (other lateral wall) – episode of care unspecified
410.51 AMI (other lateral wall) – initial episode of care
410.60 AMI (true posterior wall) – episode of care unspecified
410.61 AMI (true posterior wall) – initial episode of care
410.70 AMI (subendocardial) – episode of care unspecified
410.71 AMI (subendocardial) – initial episode of care
410.80 AMI (other specified site) – episode of care unspecified
410.81 AMI (other specified site) – initial episode of care
410.90 AMI (unspecified site) – episode of care unspecified
410.91 AMI (unspecified site) – initial episode of care

ICD-10 Codes that define the patient cohort:
I2109 ST elevation (STEMI) myocardial infarction involving other coronary artery of anterior wall
I2119 ST elevation (STEMI) myocardial infarction involving other coronary artery of inferior wall
I2111 ST elevation (STEMI) myocardial infarction involving right coronary artery
I2119 ST elevation (STEMI) myocardial infarction involving other coronary artery of inferior wall
I2129 ST elevation (STEMI) myocardial infarction involving other sites
I214 Non-ST elevation (NSTEMI) myocardial infarction
I213 ST elevation (STEMI) myocardial infarction of unspecified site

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The mortality measures exclude index admissions for patients:
1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility.
2. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data;
3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission; or
4. Discharged against medical advice (AMA).

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

For Medicare FFS patients, the measure additionally excludes admissions for patients without at least 30 days post-discharge enrollment in FFS Medicare (because the 30-day mortality outcome cannot be assessed in this group).

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
1. The discharge disposition indicator is used to identify patients alive at discharge. Transfers are identified in the claims when a patient with a qualifying admission is discharged from an acute care hospital and admitted to another acute care hospital on the same day or next day. In addition, patient length of stay and condition is identified from the admission claim.
2. Inconsistent vital status or unreliable data are identified if any of the following conditions are met 1) the patient's age is greater than 115 years; 2) if the discharge date for a hospitalization is before the admission date; and 3) if the patient has a sex other than 'male' or 'female'.
3. Hospice enrollment in the 12 months prior to or on the index admission is identified using hospice data and the Inpatient standard analytic file (SAF). This exclusion applies when the measure is used in Medicare FFS patients only.
4. Discharges against medical advice (AMA) are identified using the discharge disposition indicator.

Additional exclusions:
• AMI admissions within 30 days of discharge from a qualifying index admission, which are identified by comparing the discharge date from the index admission with the readmission date.
• Admissions without at least 30 days post-discharge enrollment in FFS Medicare, which is determined by examining the Medicare Enrollment Database (EDB)

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et. al., 2006).

The measure employs a hierarchical logistic regression model to create a hospital level 30-day RSMR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, the model adjusts the log-odds of mortality within 30-days of admission for age, sex, and selected clinical covariates. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a death at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables:
Candidate variables were patient-level risk-adjustors that were expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment including age, sex, and indicators of comorbidity and disease severity. For each patient, covariates are obtained from Medicare claims extending 12 months prior to and including the index admission. However, in the all-payer hospital discharge database measure, the risk-adjustment variables can be obtained only from inpatient claims in the prior 12 months and the index admission (this was tested explicitly in our all-payer testing, as many all-payer datasets do not include outpatient claims).

The model adjusts for case-mix differences based on the clinical status of patients at the time of admission. We used condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes, and combinations of CCs as candidate variables (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12-months prior, and not complications that arise during the course of the hospitalization, are included in the risk-adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care and that are only recorded in the index admission.

The final set of risk adjustment variables are:
Demographics
Male
Age-65 (years above 65, continuous) for 65 and over cohorts; or Age (years, continuous) for 18 and over cohorts.

Comorbidities
Congestive heart failure (CC 80)
Acute myocardial infarction (CC 81)
Other acute/subacute forms of ischemic heart disease (CC 82)
Anterior myocardial infarction (ICD-9 codes 410.00-410.19)
Other location of myocardial infarction (ICD-9 codes 410.20-410.69)
Coronary atherosclerosis or angina (CC 83, 84)
Cardio-respiratory failure and shock (CC 79)
Valvular and rheumatic heart disease (CC 86)
Hypertension (CC 89, 91)
Stroke (CC 95-96)
Cerebrovascular disease (CC 97-99, 103)

Renal failure (CC 131)
Chronic obstructive pulmonary disease (COPD) (CC 108)
Pneumonia (CC 111-113)
Diabetes mellitus (DM) or DM complications except proliferative retinopathy (CC 15-20, 120)
Protein-calorie malnutrition (CC 21)
Dementia or other specified brain disorders (CC 49, 50)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177, 178)
Vascular disease and complications (CC 104, 105)
Metastatic cancer, acute leukemia and other severe cancers (CC 7, 8)
Trauma in last year (CC 154-156, 158-162)
Major psychiatric disorders (CC 54-56)
Chronic Liver Disease (CC 25-27)
History of CABG (ICD-9-CM V45.81, 36.10-36.16)
History of PTCA (ICD-9-CM V45.82, 00.66, 36.01, 36.02, 36.05, 36.06, 36.07)

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

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

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Available at measure-specific web page URL identified in S.1.

S.16. Type of score:
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*
The measure estimates hospital-level 30-day all-cause RSMRs following hospitalization for AMI using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of mortality within 30 days of discharge using age, sex, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of mortality at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

The RSMR is calculated as the ratio of the number of "predicted" to the number of "expected" deaths, multiplied by the national unadjusted mortality rate. For each hospital, the numerator of the ratio ("predicted") is the number of deaths within 30 days

predicted on the basis of the hospital's performance with its observed case mix, and the denominator ("expected") is the number of deaths expected on the basis of the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital's performance given its case mix to an average hospital's performance with the same case mix. Thus, a lower ratio indicates lower-than-expected mortality or better quality and a higher ratio indicates higher-than-expected mortality or worse quality.

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

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed readmission rate. The hierarchical logistic regression models are described fully in the original methodology report (Krumholz et al., 2005).

References:
1. Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.
2. Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Models for AMI and HF 30-Day Mortality Methodology. 2005.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A. This measure is not based on a sample or survey.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A. This measure is not based on a sample or survey.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing values are rare among variables used from claims data in this measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Other, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Data sources for the Medicare FFS measure:
1. Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for fee-for service inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician claims for the 12 months prior to an index admission.
2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992).
3. Veterans Health Administration Data: This data source contains claims data for VA inpatient and outpatient services including: inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician claims for the 12 months prior to and including each index admission. Unlike Medicare FFS

patients, VA patients are not required to have been enrolled in Part A and Part B Medicare for the 12 months prior to the date of admission.

All-payer data sources:
For our analyses to examine use in all-payer data, we used all-payer data from California in addition to CMS data for Medicare FFS 65+ patients in California hospitals. California is a diverse state, and, with more than 37 million residents, California represents 12% of the US population. We used the California Patient Discharge Data, a large, linked database of patient hospital admissions. In 2006, there were approximately 3 million adult discharges from more than 450 non-Federal acute care hospitals. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations and to evaluate rates of both readmission and mortality (via linking with California vital statistics records).

Using all-payer data from California as well as CMS Medicare FFS data for California hospitals, we performed analyses to determine whether the AMI mortality measure can be applied to all adult patients, including not only FFS Medicare patients aged 65+ but also non-FFS Medicare patients aged 65+ and younger patients aged 18-64 years at the time of admission.

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

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0230
Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older
Date of Submission: Click here to enter a date
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23*)**	**Measure Tested with Data From:**
☐ abstracted from paper record	☒ abstracted from paper record

☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/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: Census data/American Community Survey

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

The datasets used for testing included Medicare Parts A and B claims, Veterans' Health Administration claims, as well as the Medicare Enrollment Database (EDB). Additionally, census data were used to assess socio-demographic factors. The dataset used varies by testing type; see Section 1.7 for details.

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

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

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

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

For this measure, hospitals are the measured entities. All non-federal, acute inpatient US hospitals (including territories) with Medicare fee-for-service (FFS) beneficiaries over the age of 65 are included. All Veteran's Health Administration Hospitals are also included in the current publically reported measure. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

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

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

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

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

For reliability testing (Section 2a2: Dataset 1), measure exclusions (Section 2b3; Dataset 1), and testing to identify meaningful differences in performance (Section 2b5; Dataset 1):
The reliability of the model was tested by randomly selecting 50% of the Medicare FFS patients aged 65+ in the most recent 3-year cohort and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients and compared the two. In each year of measure maintenance we also re-fit the model and compare the frequencies and model coefficients of risk variables (condition categories for patient comorbidities) and model fit across 3 years (Dataset 1 below).

Dataset 1 (current public reporting cohort): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: July 1, 2011 – June 30, 2014 (current public reporting cohort)
Number of admissions: 497,550
Number of patients in sample A = 247,641
Number of patients in sample B = 249,909
Patient Descriptive Characteristics: average age= 78.9, % male= 52.4
Number of measured entities: 4,490

For validity testing (Section 2b2; Datasets 2 and 4):
Dataset 2 (Chart validation sample data): Cooperative Cardiovascular Project (CCP) initiative medical chart-abstracted AMI cases, linked to the corresponding Medicare Part A Inpatient and Outpatient and Part B Outpatient claims, and mortality data from the Medicare enrollment database.
Dates of Data: 1994-1995
Number of Admissions: 181,032
Number of Measured Entities: 4,668

Dataset 4 (California discharge data): Medicare FFS and all-payer data
Dates of Data: January 1, 2006- December 31, 2006
Number of Admissions:
Total discharges of all payer (18+): 39,481
Total discharges of Medicare FFS 65+ patients: 11,418
Patient Descriptive Characteristics (for FFS 65+): average age=80, % male=50
Number of Measured Entities: >450 non-Federal acute care hospitals

For testing of measure risk adjustment (Section 2b4; Datasets 1, 3 and 4)
Dataset 1: Please see description above

Dataset 3 (Original development dataset from 1998): Medicare Part A Inpatient and Outpatient and Part B Outpatient claims
Dates of Data: 1998
Number of Admissions: 134, 661
Number of Measured Entities: 4,646

Dataset 4 (California discharge data): Please see description above.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We selected sociodemographic status (SDS) variables to analyze after reviewing the literature and examining available national data sources. Few patient-level SDS variables that can be linked to Medicare data are available nationally. We found both race [black vs. non-black] and dual-eligible status, e.g. enrolled in both Medicare and Medicaid, [obtained from CMS claims enrollment data] as the only two patient-level SDS variables available to examine directly. While some

argue against consideration of race in risk-adjustment, we felt it was important to understand the association with race as well as more traditional socio-economic variables.

We also considered neighborhood-level variables, linked by patient zip code level data that could serve in a risk model as a proxy for patient-level SDS. A range of census-collected SDS variables [collected annually as part of American Community Survey and aggregated over 5-years] including income and education, were available. Currently we are only able to link the data at a 5-digit zip code level. Nine-digit zip code data may provide a more granular view of patient sociodemographic status, but this data is not available to us at this time and we therefore cannot ascertain the incremental, if any, value of greater geographic discrimination for risk adjustment purposes.

Our conceptual model and the literature regarding how SDS may influence post-discharge mortality did not identify a single SDS concept as predominant in the pathway. There is a large body of literature linking various SDS factors to worse health status and higher mortality over a lifetime (Adler and Newman 2002, Mackenbach et al. 2000, Tonne et al. 2005, van Oeffelen et al. 2012). Income, education, and occupational level are the most commonly examined variables. However, literature directly examining how different SDS factors might influence the likelihood of mortality in older, insured, Medicare patients within 30 days of an admission for cardiovascular disease is much more limited. Assuming that the risk imparted based on zip code level data may reflect multiple different SDS variables, we chose to analyze a validated AHRQ composite index of SDS, which has been used and tested among Medicare beneficiaries (Blum et al. 2014; Bonito et al. 2008). This index is a composite of 7 different variables found in the census data which may capture SDS better than any single variable. The index variables include rates of unemployment, percent of person living below poverty, education level (percent below 12^{th} grade education and percent with college education), crowding (average of more than one person per room), median household income, and median housing value. We identified patients as low SDS if they lived in a neighborhood in the lowest quartile of this index.

Other variables can be found at a county or regional level and could represent the hospital's community. We did not directly test any such variables because they are not as closely related to patients' sociodemographic status given the wide scope of a county and seemed unlikely to be ideal for patient-level risk adjustment.

References:

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

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

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

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

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

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

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*)

Data Element Reliability

In constructing the measure, we aim to utilize only those data elements from the claims that have both face validity and reliability. We avoid the use of fields that are thought to be coded inconsistently across hospitals or providers. Specifically, we use fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of CMS auditing and billing policies while seeking to avoid variables that do not meet this standard. For example, "discharge disposition" is a variable in Medicare claims data that is not thought to be a reliable variable for identifying a transfer between two acute care facilities. Thus, we derive a variable using admission and discharge dates as a surrogate for "discharge disposition" to identify hospital admissions involving transfers. This allows us to identify these admissions using variables in the claims data which have greater reliability than the "discharge disposition" variable. In addition, CMS has several hospital auditing programs used to assess overall claims code accuracy in order to ensure appropriate billing and assist with overpayment recoupment. CMS routinely conducts data analyses to identify potential problem areas and detect fraud. Audits are important data fields used in our measures including diagnosis, procedure codes, and other elements that are consequential to payment.

Finally, we assess the reliability of the data elements by comparing model variable frequencies and odds ratios from logistic regression models across three years of data.

Measure Score reliability

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

For test-retest reliability, we combined index admissions from successive measurement periods into one dataset, randomly sampled half of the patients within each hospital, calculated the measure for each hospital, and repeated the calculation using the second half. Thus, each hospital is measured twice but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, we calculated the intra-class correlation coefficient (ICC) (Shrout and Fleiss, 1979), and assessed the values according to conventional standards (Landis and Koch, 1977). Specifically, we used Dataset 1 to randomly split the samples, and calculated the RSMR for each hospital within each sample. The agreement of the two RSMRs was quantified for hospitals in each sample using the intra-class correlation as defined by ICC (2,1) by Shrout and Fleiss (1979).

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

References:

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

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-428.

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

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

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

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

Data element reliability results (Dataset 1)

The frequency of some model variables increased. The increase may reflect an increased rate of comorbidities in the fee-for-service population but is also due, in part, to increased hospital coding of comorbidities. In the 2012 update to the measures, we increased the number of diagnosis codes and procedure codes to align with the Version 5010 format changes required by the Department of Health and Human Services (DHHS). Beginning in 2010, hospitals could submit up to 25 diagnosis and procedure codes. Over time (from July 2011-June 2012 to July 2013-June2014), more hospitals have submitted increased numbers of codes, which translate into increased frequencies for some model variables. Some notable decreases occurred in congestive heart failure (CC 80) (31.0% to 29.3%), pneumonia (CC 111-113) (23.7% to 22.3%), and dementia or other specific brain disorders (CC 49-50) (20.8% to 19.8%), while notable increases occurred in history of percutaneous transluminal coronary angioplasty (PTCA) (15.9% to 17.5%), renal failure (CC 131) (26.3% to 27.3%), and % males (51.7% to 53.3%).

Measure Score Reliability Results (Dataset 1)

There were 497,550 admissions in the combined three-year sample, with 247,641 patients in one sample and 249,909 patients in the other randomly selected sample. The agreement between the two RSMRs for each hospital was 0.41, which according to the conventional interpretation is “moderate” (Landis & Koch, 1977).

Note that this analysis was limited to hospitals with 12 or more cases in each split sample. The intra-class correlation coefficient is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is reported with the full three years of data. The correlation coefficient is expected to be higher using the full three-year sample since it would include more patients.

Reference:

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

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

The stability over time of the risk factor odds ratios suggests that the underlying data elements are reliable. Additionally, the ICC score demonstrates moderate agreement across samples using a conservative approach to assessment.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

During original measure development, we validated the AMI mortality administrative model against a medical record model in the same cohort of patients for which hospital-level AMI mortality medical record data were available.

For the derivation of the chart-based model, we used cases identified through the Cooperative Cardiovascular Project (CCP) initiative and provided by the Health Care Financing Administration (now CMS). The CCP initiative included more than 200,000 admissions to non-governmental, acute care hospitals in the United States and Puerto Rico (Krumholz et al., 1998; Marciniak et al., 1998). In the CCP study, CMS sampled all claims from FFS Medicare patients during an approximately 8-month period (varying by state) in 1994 and 1995 who were discharged with a principal diagnosis of AMI (ICD-9-CM code 410, excluding 410.x2). These patients were matched to the Medicare enrollment database to determine survival and, where applicable, the date of death. Corresponding medical records were abstracted by 2 clinical data abstraction centers (DynKePRO [York, PA] and FMAS Corporation [Rockville, MD]), and the clinical data used to confirm the diagnosis of AMI.

The test sample contained 178,188 cases with an unadjusted mortality rate of 19.0%.

The medical record model validation included clinician and hospital outpatient data. The same coding and transfer rules described in the AMI administrative dataset were used in defining the AMI medical record dataset.

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

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

Krumholz HM, Radford MJ, Wang Y et al. National use and effectiveness of beta-blockers for the treatment of elderly patients after acute myocardial infarction: National Cooperative Cardiovascular Project. *JAMA.* 1998;280:623-629.

Marciniak TA, Ellerbeck EF, Radford MJ et al. Improving the quality of care for Medicare patients with acute myocardial infarction: results from the Cooperative Cardiovascular Project. *JAMA.* 1998;279:1351-1357.

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

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary

Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were initially identified using 2015 General Equivalence Mapping (GEM) software. Clinicians with expertise in relevant areas were enlisted to select and evaluate which ICD-10 codes map to the ICD-9 codes currently in use for this measure. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Dataset 2 (see section 1.7 above):
The performance of the administrative and medical record models is similar. The areas under the receiver operating characteristic (ROC) curve are 0.69 and 0.77, respectively, for the two models. In addition, they are similar with respect to predictive ability. For the administrative model, the predicted readmission rate ranges from 4.79%, in the lowest predicted decile, to 39.33%, in the highest predicted decile, with a range of 36%. For the medical record model, the corresponding range is 2.9% to 57.7%with a range of 54.8%.

We estimated hospital-level RSMRs using the corresponding hierarchical logistic regression administrative and medical record models for the linked patient sample. We then examined the linear relationship between the two sets of estimates using regression techniques and weighting by the total number of cases in each hospital. The correlation coefficient of the standardized rates from the administrative and medical record models is 0.91.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)

The results between the administrative and medical record models proved to be similar in each of the model testing that was performed. The ROC results were similar and in line with other mortality models. The correlation between the resulting RSMRs calculated from both models was 0.91 which shows that there was a strong correlation in rates calculated from the clinical and admin models.

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion. These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

2b3.2. 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*)

Among 4,490 hospitals with at least 25 index stays in [July 1, 2011- June 30, 2014] (Dataset 1):

Exclusion	N	%	Distribution across hospitals: Min, 25th, 50th, 75th percentile, max
1. Discharged alive on the day of admission or the following day who were not transferred to another acute care facility.	34,657	6.17%	Min: 0.00% 25th: 0.00% 50th: 4.04% 75th: 8.20% Max: 100.00%
2. With inconsistent or unknown vital status or other unreliable demographic (age and gender) data.	26	<0.01%	Min: 0.00% 25th: 0.00% 50th: 0.00% 75th: 0.00% Max: 4.35%
3. Enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission, including the first day of the index admission.	4,770	0.85%	Min: 0.00% 25th: 0.00% 50th: 0.00% 75th: 1.13% Max: 100.00%
4. Discharged against medical advice (AMA).	2,630	0.47%	Min: 0.00% 25th: 0.00% 50th: 0.00% 75th: 0.34% Max: 100.00%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusion 1 (patients who were discharged alive on the day of admission or the following day who were not transferred to another acute care facility) is meant to ensure a clinically coherent cohort (excludes 6.2% of cohort). This exclusion prevents the inclusion of patients who likely did not suffer a clinically significant AMI. For most hospitals, this results in very few patients being excluded. For those hospitals with greater proportions of excluded patients, the measure is likely excluding less severe patients that may not be considered as AMI at other hospitals. This exclusion was guided by the input of clinical experts at the time of measure development. Exclusion 2 (patients with inconsistent or unknown vital status or other unreliable demographic (age and gender) data) and Exclusion 3 (patients enrolled in the Medicare hospice program or used VA hospice services any time in the 12 months prior to the index admission), including the first day of the index admission, are necessary for valid calculation of the measure. Exclusion 4 (patients who leave AMA) is needed for acceptability of the measure to hospitals who do not have the opportunity to adequately prepare such patients for discharge. The aforementioned exclusions are unlikely to affect the measure score because they exclude a very small percentage of patients.

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

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 27 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Our approach to risk adjustment was tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al. 2006).

The measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital level 30-day RSMR. This approach to modeling appropriately accounts for the structure of the data (patients clustered within hospitals), the underlying risk due to patients' comorbidities, and the sample size at a given hospital when estimating hospital mortality rates. In brief, the approach simultaneously models two levels (patient and

hospital) to account for the variance in patient outcomes within and between hospitals (Normand and Shahian et al. 2007). At the patient level, each model adjusts the log-odds of mortality within 30 days of admission for age, sex, selected clinical covariates, and a hospital specific intercept. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept, or hospital specific effect, represents the hospital contribution to the risk of mortality, after accounting for patient risk and sample size, and can be inferred as a measure of quality. The hospital-specific intercepts are given a distribution in order to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, the hospital intercepts should be identical across all hospitals after adjusting for patient risk.

Clinical Factors

Candidate and Final Risk-adjustment Variables:

The measure was developed using Medicare FFS claims data. Candidate variables were patient-level risk-adjustors that are expected to be predictive of mortality based on empirical analysis, prior literature, and clinical judgment including demographic factors (age, sex) and indicators of comorbidity and disease severity. For each patient, covariates were obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusted for case differences based on the clinical status of the patient at the time of admission. We used condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes. We did not risk-adjust for CCs that were possible adverse events of care and that were only recorded in the index admission. In addition, only comorbidities that conveyed information about the patient at that time or in the 12 months prior, and not complications that arose during the course of the hospitalization, were included in the risk-adjustment.

The final set of risk-adjustment variables is:

Demographic

- Age-65 (years above 65, continuous) for 65 and over cohorts; or Age (years, continuous) for 18 and over cohorts.
- Male

Cardiovascular

- History of PTCA
- History of CABG
- Congestive heart failure
- History of AMI
- Unstable angina
- Anterior myocardial infarction
- Other location of myocardial infarction
- Chronic atherosclerosis
- Cardio-respiratory failure and shock
- Valvular and rheumatic heart disease

Comorbidity

- Hypertension
- Stroke
- Cerebrovascular disease
- Renal failure
- Chronic Obstructive Pulmonary Disease
- Pneumonia
- Diabetes and DM complications
- Protein-calorie malnutrition
- Dementia and senility

• Hemiplegia, paraplegia, paralysis, functional disability
• Peripheral vascular disease
• Metastatic cancer, acute leukemia and other severe cancers
• Trauma in the last year
• Major psychiatric disorders
• Chronic liver disease

Sociodemographic Factors

We selected candidate sociodemographic factors for examination based on a review of the literature, conceptual pathways, and feasibility. In section 1.8, we describe the variables that we considered and analyzed based on this review. Below we describe the pathways by which SDS may influence 30-day mortality.

Our conceptualization of the pathways by which patient SDS affects 30-day mortality is informed by the literature. Although there is a long list of studies showing a relationship between lower SDS status and mortality rates generally, there is relatively little literature directly examining how SDS might influence the likelihood of older, insured, Medicare patients dying within 30 days of a admission for cardiovascular disease and even less literature to directly illuminate the pathways by which SDS influences this outcome. The ability to distinguish between these pathways is challenging but important for making decisions regarding risk adjustment.

One important pathway by which patient SDS influences 30-day mortality is through health status at the time of admission. SDS factors can influence admission health status both due to the impact of multiple related stressors over a lifetime contributing to overall worse health as well as through poor access to care and potentially delayed presentation. This results in low SDS patients, when compared with other patients, often arriving for hospital admission with greater levels of illness or comorbidity burden. This pathway should be largely accounted for in our current clinical risk-adjustment.

However, there are a number of other pathways by which patient SDS may influence 30-day mortality that are related to hospital quality. The first, sometimes referred to as contextual effects, is that patients of low SDS may present at lower quality institutions for care. Therefore, some part of the apparent relationship between SDS and mortality may be due to clustering of patients of low SDS at poorer quality institutions (Barnato et al. 2005; Hasnain-Wynia et al. 2010; Jha et al. 2011; Skinner et al. 2005).

Next, within the hospital, patients of low SDS may receive differentiated care as compared to counterparts of higher SDS – this can occur for a variety of reasons and in some cases differentiated care could be worse quality and in others better quality. For example, providers may be less likely to offer guideline-concordant care to patients of low SDS – perhaps based on discrimination or misunderstanding of patients' wishes and values (Institute of Medicine, 2009). However, in other cases, differentiated care may be that patients of low SDS appropriately need different types of care or services such as low literacy information, social worker support or transportation at discharge. Providing needed differentiated care is patient-centered and appropriate – the equivalent to ensuring a diabetic goes home with insulin – however, low SDS patients may not always receive needed differentiated care. This lack of needed differentiated care may also contribute to relatively worse outcomes.

Finally, there may be pathways by which SDS influences 30-day mortality risk outside of health care quality and admission health status. Some SDS factors may affect the likelihood of mortality without directly affecting health status on admission or the quality of care received during the hospital stay. For instance, despite a hospital making appropriate care decisions and providing tailored care and education, a lower-income patient may elect not to follow prescribed care (e.g. refill a prescription or keep a follow-up visit with a primary care provider) because limited resources create competing priorities for the patient.

These sets of proposed pathways are complex to distinguish analytically. We, therefore, first assessed if there was evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways. First we evaluated the variation in the prevalence of low SDS patients among providers. We then assessed the relationship

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

Based on this model and the considerations outlines in 1.8, the following SDS variables were considered:

- Dual eligible status
- African American race
- AHRQ SES index

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

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

IOM (Institute of Medicine). 2009. Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. Washington, DC: The National Academies Press.
Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and medicaid patients. Health affairs (Project Hope). 2011;30(10):1904-1911.

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

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

Skinner J, Chandra A, Staiger D, Lee J, McClellan M. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation. 2005;112(17):2634-2641.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

A table of candidate variables and odds ratios associated with each variable in the model are available in the attached Excel file (referenced in data field S.2b) (Dataset 1).

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Variation in prevalence of the factor across measured entities

The prevalence of SDS factors in the AMI cohort varies substantially across hospitals. The median percent of dual eligible patients is 10.8% (interquartile range [IQR] 6.9%-16.8%). The median percentage of black patients is 3.6% (IQR 0.7%-10.6%). The median frequency of low SES AHRQ indicator patients is 16.4% (IQR 4.1%-40.3%).

Empirical association with the outcome (bivariate)

The patient-level observed AMI unadjusted mortality rate for dual-eligible patients was somewhat higher, at 16.1% compared with 14.0% for all other patients. The mortality rate for black patients was lower at 12.6% compared with 14.4% for patients of all other races. The mortality rate for patients in the lowest SES quartile by AHRQ Index was slightly higher at 14.4% compared with 13.9% for patients in the highest SES quartile.

Incremental effect of SDS variables in a multivariable model

We then examined the strength and significance of the SDS variables in the context of a multivariable model. Each of the variables remained significantly associated in the multivariable model.

For dual eligibility and the AHRQ SES indicator, the variable is associated with higher risk of modest strength. Odds ratios are on the order of 1.12 for dual eligibility and 1.09 for AHRQ SES. This is similar to the odds ratio for comorbidities such as COPD and substantially lower than the risk associated with comorbidities such as metastatic cancer. For race, black race is associated with a lower risk of mortality, with an odds ratio of 0.81. In all cases, the c-statistic for the AMI patient-level multivariate model with the SDS variable in the model is essentially unchanged from that without.

To further understand the relative importance of these risk-factors in the measure we compared hospital performance with and without the addition of each SDS variable. We found that the addition of any of these variables into the model has little to no effect on hospital performance. The mean absolute change in hospitals' RSMRs when adding a dual eligibility indicator is -0.00039% with a correlation coefficient between RSMRs for each hospital with and without dual eligibility added of 0.9996. The mean absolute change in hospitals' RSMRs when adding a race indicator is 0.00087% with a correlation coefficient between RSMRs for each hospital with and without race added of 0.9971. The mean absolute change in hospitals' RSMRs when adding a low SES AHRQ indicator is -0.00205% with a correlation coefficient between RSMRs for each hospital with and without low SES added of 0.9982.

Overall we found that among the SDS variables that could be feasibly incorporated into this model, 1) the relationship with mortality is small and 2) the relative effect of black race is stronger than the other two SDS variables in the opposite direction than what has been the expressed concern of stakeholders interested in adding such adjustment to the models. We also found that the impact of adding any of these indicators is very small to negligible on model performance and hospital profiling.

Given these findings and complex pathways that could explain any relationship between SDS and mortality, which do not all support risk-adjustment, we did not incorporate SDS variables into the measure.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AGE_65	1.059	1.058	1.060
MALE	1.151	1.131	1.171
HXPCI	0.751	0.731	0.770
HXCABG	1.088	1.060	1.117
AMI_ANT	2.223	2.160	2.287
AMI_OTH	1.671	1.628	1.716
Hx_CHF	1.320	1.293	1.348
Hx_MI	0.970	0.944	0.997
UnAngina	0.921	0.895	0.947
Atherosc	0.607	0.594	0.621
RespFail	1.184	1.153	1.216
ValvuDis	1.084	1.064	1.104
HTN	0.723	0.704	0.743
Stroke	1.026	0.993	1.059
CerebDis	0.961	0.940	0.982
RenFail	1.199	1.175	1.224
COPD	1.122	1.101	1.144
Pneumon	1.533	1.503	1.564

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Diabetes	1.096	1.077	1.115
PCMalnut	1.641	1.595	1.687
Dementia	1.439	1.410	1.468
FunctDis	1.216	1.177	1.256
PVDis	1.081	1.060	1.102
MetasCA	2.007	1.937	2.079
Trauma	0.994	0.977	1.012
PsychDis	1.080	1.049	1.111
LiverDis	1.512	1.423	1.606
dual_elig	1.122	1.094	1.151

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AGE_65	1.058	1.057	1.059
MALE	1.144	1.124	1.164
HXPCI	0.748	0.729	0.767
HXCABG	1.084	1.055	1.113
AMI_ANT	2.217	2.155	2.281
AMI_OTH	1.665	1.622	1.710
Hx_CHF	1.326	1.299	1.354
Hx_MI	0.971	0.945	0.998
UnAngina	0.922	0.896	0.948
Atherosc	0.605	0.592	0.619
RespFail	1.182	1.151	1.214
ValvuDis	1.080	1.061	1.100
HTN	0.726	0.707	0.745
Stroke	1.028	0.995	1.061
CerebDis	0.959	0.938	0.980
RenFail	1.205	1.181	1.229
COPD	1.127	1.105	1.148
Pneumon	1.536	1.506	1.567
Diabetes	1.102	1.083	1.121
PCMalnut	1.651	1.606	1.698
Dementia	1.452	1.423	1.481
FunctDis	1.228	1.189	1.268
PVDis	1.084	1.063	1.105
MetasCA	1.998	1.928	2.070
Trauma	0.992	0.974	1.010
PsychDis	1.086	1.055	1.118

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
LiverDis	1.524	1.435	1.619
black	0.810	0.767	0.855

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AGE_65	1.059	1.058	1.060
MALE	1.148	1.129	1.168
HXPCI	0.750	0.731	0.770
HXCABG	1.087	1.059	1.117
AMI_ANT	2.224	2.162	2.288
AMI_OTH	1.672	1.629	1.717
Hx_CHF	1.321	1.294	1.349
Hx_MI	0.969	0.943	0.996
UnAngina	0.921	0.895	0.947
Atherosc	0.606	0.593	0.620
RespFail	1.185	1.154	1.217
ValvuDis	1.084	1.065	1.104
HTN	0.723	0.704	0.742
Stroke	1.026	0.994	1.060
CerebDis	0.959	0.939	0.981
RenFail	1.200	1.176	1.225
COPD	1.123	1.102	1.144
Pneumon	1.535	1.505	1.566
Diabetes	1.097	1.078	1.116
PCMalnut	1.640	1.595	1.686
Dementia	1.444	1.416	1.473
FunctDis	1.221	1.182	1.262
PVDis	1.084	1.063	1.105
MetasCA	2.006	1.936	2.078
Trauma	0.996	0.978	1.014
PsychDis	1.087	1.057	1.119
LiverDis	1.517	1.428	1.611
group_sesind	1.094	1.073	1.115

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model <u>or</u> stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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

If stratified, skip to 2b4.9

Approach to assessing model performance

During measure development, we computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the development and validation cohort:

Discrimination Statistics:

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

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

Calibration Statistics:

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

We tested the performance of the model in all four datasets described in section 1.7.

Citation

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

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

For the development cohort the results are summarized below (Dataset 3):

1st half of randomly split development sample: C statistic = 0.71

Predictive ability (lowest decile %, highest decile %) = (4.0%, 40.0%)

2st half of randomly split development sample: C statistic = 0.70

Predictive ability (lowest decile %, highest decile %) = (4.2%, 40.1%)

For the current measure cohort the results are summarized below (Dataset 1):

C statistic = 0.72

Predictive ability (lowest decile %, highest decile %) = (2.8%, 33.3%)

For comparision of model with and without inclusion of SDS factors see above section.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

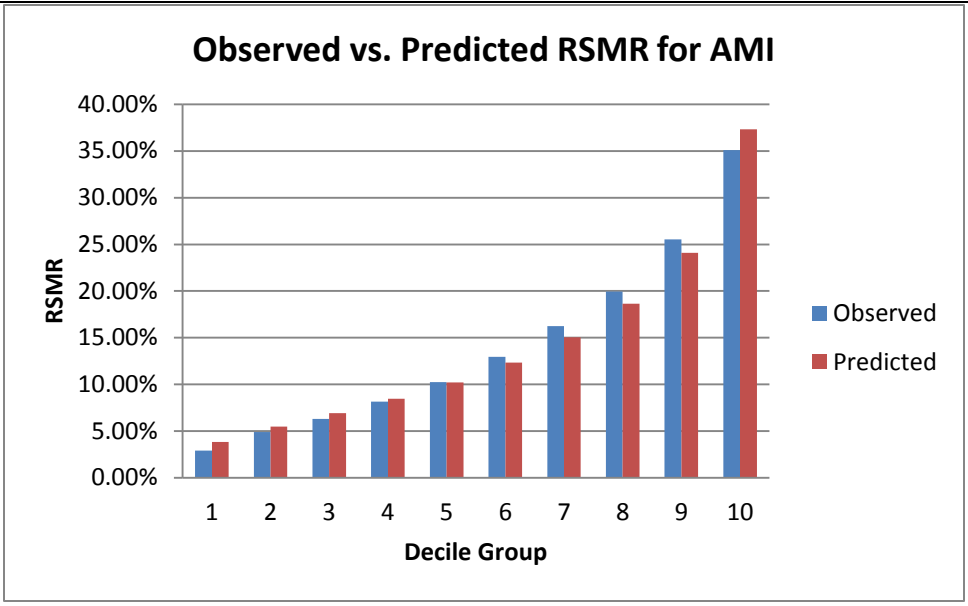
Development spilt dataset (Dataset 3):

1st half of randomly split development sample: Calibration: (0.000, 1.000)

1st half of randomly split development sample: Calibration: (-0.030, 0.994)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

The risk decile plot is a graphical depiction of the deciles calculated to measure predictive ability. Below, we present the risk decile plot showing the distributions for the [Observed vs. Predicted RSMR for AMI] for Medicare FFS data from July 2011 to June 2014 (Dataset 1).



Plots for other datasets were similar.

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Discrimination Statistics
The C-statistic of 0.72 indicates good model discrimination. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

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

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

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

2b4.11. 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*)

Application to Medicare FFS Beneficiaries Using Inpatient Data Only for Risk Adjustment

As part of the testing for the model in the all-payer data, we also applied the model to CMS data for Medicare FFS 65+ patients in California hospitals using only inpatient data for risk adjustment. Specifically, we created a 2006 measure cohort with complete one-year history data and 30-day follow-up data (N= 39,481).

To help determine whether the measure could be applied to Medicare FFS 65+ patients using only Medicare Part A data, we performed analyses to assess how the model performs when using only inpatient claims data for risk adjustment, as all-payer hospital discharge databases do not have outpatient claims. To assess the validity of using only admission claims data for risk adjustment, we fit the model separately using the full dataset and using only admission claims data and (a) compared the odds ratios (ORs) for the various risk factors; (b) conducted a reclassification analysis to compare risk prediction at the patient level; (c) compared model performance in terms of the c-statistic (discrimination); and (d) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and R2) to assess whether the model with only admission claims data is different from the current model in profiling hospital rates.

Analyzing CMS data for Medicare FFS 65+ beneficiaries in California hospitals: (a) the magnitude of odds ratios for most risk factors was similar when comparing the model using the full dataset and using only admission claims data; (b) when comparing the model with the full dataset and with only admission claims data, the reclassification analysis demonstrated good patient-level risk prediction; (c) the c-statistic was similar (0.713 vs. 0.725); and (d) hospital-level risk-standardized rates were highly correlated (ICC=0.984).

Application to Patients Aged 18 and Older

We also applied the model to all-payer data from California. The analytic sample included 39,481 cases aged 18 and older in the 2006 California Patient Discharge Data. When used in all-payer data, only admission claims data are used for risk adjustment, as the hospital discharge databases do not have outpatient claims.

To help determine whether the measure could be applied to a population of patients aged 18+, we examined the interaction terms between age (18-64 vs. 65+) and each of the other risk factors. Specifically, we fit the model in all patients 18+ with and without interaction terms and (a) conducted a reclassification analysis to compare risk prediction at the patient level; (b) compared the c-statistic; and (c) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and R-square) to assess whether the model with interactions is different from the current model in profiling hospital rates.

When the model was applied to all patients 18+, overall discrimination was good (c-statistic=0.765). In addition, there was good discrimination and predictive ability in both those aged 18-64 and those aged 65+. Moreover, the distribution of Pearson residuals was comparable across the patient subgroups. When comparing the model with and without interaction terms, (a) the reclassification analysis demonstrated that nearly all patients were found to be in a similar risk category; (b) the c-statistic was nearly identical (0.767 vs. 0.765); and (c) hospital-level risk-standardized rates were highly correlated (ICC=0.998). Thus, the inclusion of the interactions did not substantively affect either patient-level model performance or hospital-level results.

Therefore, the measure can be applied to all-payer data for patients 18 and older.

Reference:

Harrell FE, Shih YCT. Using full probability models to compute probabilities of actual interest to decision makers. Int J Technol Assess Health Care. 2001;17:17–26.

Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (YNHHSC/CORE) (January 2012). Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following Pneumonia Hospitalization. In *Testing Publicly Report 30-Day Acute Myocardial Infarction, Heart Failure, and Pneumonia Risk-Standardized Mortality and Readmission Measures in California All-Payer Data.*

Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (YNHHSC/CORE) (January 2012). Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following Chronic Obstructive Pulmonary

Disease (COPD) Hospitalization. In *Testing Chronic Obstructive Pulmonary Disease 30-Day Mortality and Readmission Measures in California All-Payer Data*.

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

For public reporting of the measure, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national observed AMI mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare website as "better than the U.S. national rate" or "worse than the U.S. national rate." If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or "the difference is uncertain." CMS does not classify performance for hospitals that have fewer than 25 cases in the three-year period.

2b5.2. 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*)

Recent analyses of Medicare FFS data show substantial variation in RSMRs among hospitals. Using data from July 2011-June 2014 (Dataset 1) the mean hospital RSMR was 14.3%, with a range of 9.9% to 20.6%. The interquartile range was 13.8%-14.8%.

Of 4,490 hospitals in the study cohort, 41 performed "better than the U.S. national rate," 2,474 performed "no different from the U.S. national rate," 21 performed "worse than the U.S. national rate," and 1,954 were classified as "number of cases too small" (fewer than 25) to reliably tell how well the hospital is performing.

2b5.3. 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?*)

Despite recent decreases in mortality rates nationally, the mortality rate for AMI remains high at 13.3%.

The variation in rates and number of performance outliers suggests there are differences in the quality of care received across hospitals for AMI mortality that support measurement to reduce the variation.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note*: This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

2b7.2. 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*)

2b7.3. 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*)

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Administrative data are routinely collected as part of the billing process.

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)*.

There are no fees associated with the use of this measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. **Current and Planned Use**

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

Planned	Current Use (for current use provide URL)
	Public Reporting Hospital Inpatient Quality Reporting (IQR) Program http://cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

	Payment Program Hospital Value Based Purchasing (HVBP) Program http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html

4a.1. For each CURRENT use, checked above, provide:

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

Public Reporting

1. Program Name, Sponsor: Hospital Inpatient Quality Reporting (IQR) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was originally mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. Initially, the MMA provided for a 0.4 percentage point reduction in the annual market basket (the measure of inflation in costs of goods and services used by hospitals in treating Medicare patients) update for hospitals that did not successfully report. The Deficit Reduction Act of 2005 increased that reduction to 2.0 percentage points.

In addition to giving hospitals a financial incentive to report the quality of their services, the hospital reporting program provides CMS with data to help consumers make more informed decisions about their health care. Some of the hospital quality of care information gathered through the program is available to consumers on the Hospital Compare website at: http://www.medicare.gov/hospitalcompare/search.html.

Geographic area and number and percentage of accountable entities and patients included: The IQR program includes all non-federal acute care hospitals and VA hospitals in the United States. The number and percentage of accountable entities included in the program, as well as the number of patients included in the measure, varies by reporting year. For 2015 public reporting, the RSMR will be reported for 4,490 hospitals across the US. The final index cohort includes 497,550 admissions.

2. Program Name, Sponsor: Hospital Value-Based Purchasing (HVBP) Program, Centers for Medicare and Medicaid Services (CMS)

Purpose: The Hospital Value-Based Purchasing (VBP) Program is a CMS initiative that rewards acute-care hospitals with incentive payments for the quality of care they provide to people with Medicare. It was established by the Affordable Care Act of 2010 (ACA), which added Section 1886(o) to the Social Security Act. The law requires the Secretary of the Department of Health and Human Services (HHS) to establish a value-based purchasing program for inpatient hospitals. To improve quality, the ACA builds on earlier legislation—the 2003 MMA and the 2005 Deficit Reduction Act. These earlier laws established a way for Medicare to pay hospitals for reporting on quality measures, a necessary step in the process of paying for quality rather than quantity.

Geographic area and number and percentage of accountable entities and patients included: More than 3,000 hospitals across the country are eligible to participate in Hospital VBP. The program applies to subsection (d) hospitals located in the 50 states and the District of Columbia and acute-care hospitals in Maryland. Hospital VBP is based on data collected through the IQR Program.

The following hospitals are excluded from Hospital VBP:
• Hospitals and hospital units excluded from the Inpatient Prospective Payment System, such as psychiatric, rehabilitation, long-term care, children's, and cancer hospitals;
• Hospitals that do not participate in Hospital IQR during the Hospital VBP performance period;
• Hospitals cited by the Secretary of HHS for deficiencies during the performance period that pose an immediate jeopardy to patients' health or safety; and
• Hospitals that do not meet the minimum number of cases, measures, or surveys required by Hospital VBP.

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

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for

implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A, this measure is currently publicly reported

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

There has been significant progress in 30-day RSMR for AMI. The median 30-day RSMR decreased by 1.4% from 2011-2012 (median RSMR: 14.7%) to 2013-2014 (median RSMR: 13.3%). The median hospital RSMR from 2011-2014 was 14.3% (Interquartile Range [IQR] 13.8% - 14.8%). In addition, hospitals with a high proportion of Medicaid and African American patients achieve a similar range of performance as compared with hospitals with a low proportion of these patients, indicating that both groups of hospitals can perform well on the measure.

4b.2. **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.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We did not identify any unintended consequences during measure development or model testing. However, we are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative unintended consequences for patients.

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)**
0229 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older
0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
0468 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization

0505 : Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.
0506 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
1551 : Hospital-level 30-day, all-cause risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
1891 : Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization
1893 : Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization
2431 : Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI)

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

5a. Harmonization

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 completely harmonized?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
We did not include in our list of related measures any non-outcome (e.g., process) measures with the same target population as our measure. Our measure cohort was heavily vetted by clinical experts. Additionally, the measure, with the specified cohort, has been publicly reported since 2008. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure).

5b. Competing Measures

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

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

Appendix

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

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
Co.2 **Point of Contact:** Lein, Han, Lein.han@cms.hhs.gov, 410-786-0205-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)
Co.4 **Point of Contact:** Lisa, Suter, Lisa.suter@yale.edu, 203-737-3400-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org.

Our measure development team consisted of the following members:

Kanchana R. Bhat, M.P.H., Project Coordinator
Elizabeth E. Drye, M.D., S.M., Project Director
Harlan M. Krumholz, M.D., S.M., Principal Investigator
Sharon-Lise T. Normand, Ph.D., Co-Investigator*
Geoffrey C. Schreiner, B.S., Research Assistant
Yongfei Wang, M.S., Senior Statistical Analyst
Yun Wang, Ph.D., Senior Biostatistician

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2008
Ad.3 **Month and Year of most recent revision:** 04, 2015
Ad.4 **What is your frequency for review/update of this measure?** This measure is updated annually.
Ad.5 **When is the next scheduled review/update for this measure?** 04, 2016

Ad.6 **Copyright statement:** N/A
Ad.7 **Disclaimers:** N/A

Ad.8 **Additional Information/Comments:** N/A

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 #: 0669
De.2. **Measure Title:** Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac, Low Risk Surgery
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services
De.3. **Brief Description of Measure:** This measure calculates the percentage of stress echocardiography, single photon emission computed tomography myocardial perfusion imaging (SPECT MPI), or stress magnetic resonance (MR) imaging studies performed at each facility in the 30 days prior to an ambulatory non-cardiac, low-risk surgery performed at any location. The measure is calculated based on a one-year window of Medicare claims data. The measure has been publicly reported, annually, by the Centers for Medicare & Medicaid Services (CMS), since 2011, as a component of its Hospital Outpatient Quality Reporting (HOQR) Program.
1b.1. **Developer Rationale:** This measure aims to reduce overuse of cardiac imaging prior to low-risk non-cardiac surgeries, for patients with low or moderate cardiac risk, as cardiac imaging in this population can result in increased exposure to radiation with little or no clinical benefit. The measure score will guide patient selection of providers, assess quality, and inform quality improvement.

S.4. **Numerator Statement:** The number of stress echocardiography, SPECT MPI, and stress MR studies performed in a hospital outpatient department within 30 days of an ambulatory non-cardiac, low-risk surgery performed at any location (e.g., same hospital, other hospital, or physician office).
S.7. **Denominator Statement:** The number of stress echocardiography, SPECT MPI, and stress MR studies performed in a hospital outpatient department on Medicare beneficiaries within a 12-month time window.
S.10. **Denominator Exclusions:** Studies are excluded for any patients with diagnosis codes in at least three of the following categories: diabetes mellitus, renal insufficiency, stroke or transient ischemic attack, prior heart failure, or ischemic heart disease.

De.1. **Measure Type:** Process.
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility, Population : National, Population : State

IF Endorsement Maintenance – Original Endorsement Date: Apr 26, 2011 **Most Recent Endorsement Date:** Apr 26, 2011
Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: 4/26/11 **Most Recent Endorsement Date:** 4/26/11

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations are available from Standards for Imaging Efficiency: A Consensus Report for Outpatient Imaging (See Measure IEP-010-10 on Page 14)

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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. Per NQF's submission criteria, this appropriate use criteria (AUC) measure is a *process* measure, rather than an efficiency measure as defined by the developer.

The developer provides the following evidence for this process measure:

- This is a facility level measure calculates the % of stress echocardiography, single photon emission computed tomography myocardial perfusion imaging (SPECT MPI), or stress magnetic resonance (MR) imaging studies performed at each facility in the 30 days prior to an ambulatory non-cardiac, low-risk surgery performed at any location for low-risk patients.
- The developer provides the path from identifying low-risk non-cardiac surgical patients to reduced exposure to radiation, contrast agents, and more efficient use of imaging resources.
- The developer provides 2 separate guidelines with 9 guideline statements for the recommendation that patients undergoing low-risk, non-cardiac surgery should not have stress image testing. Guideline 1 recommendations ranged from Class I through Class III, Level C. Guideline 2 recommendations ranged from Class IIa through Class III, Level B & C.
- The developers provided a summary of the quantity, quality, and consistency of the evidence for the ACC/AHA guideline., and provides an additional 14 articles that support the measure's intent.

Questions for the Committee:

- *Questions specific to the measure information provided on evidence*
- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- The developer provides an analysis of Medicare fee-for-service data for 1,953 facilities for 2009-2014.
  - The maximum performance rates ranged from 14.5% in 2009 to 18.0% July 2013 through June 2014 (weighted mean = 5.07%).
  - The mean performance ranged from 5.1% in 2009 to 5.0% from July 2013 through June 2014.
- The developer states that one of the intentions of this measure is to identify facilities with significant outlying performance as demonstrated in the table.
- Using 2013 data, the developers found that race/ethnicity and facility characteristics such as size and location (urban vs. rural) had a statistically significant effect on the likelihood that a patient underwent inappropriate preoperative imaging.
- The developer also provides additional evidence from the literature demonstrating differences in the use of cardiac imaging prior to surgery based on sex, race, and age.
- Despite statistically significant differences in imaging prior to low-risk, non-cardiac surgery based on race and facility, the developer believes these differences are driven by a variation in provider practice, therefore, risk adjustment/stratification is not necessary or appropriate.

Questions for the Committee:

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

○ *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

○ *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

○ *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - There is evidence to support this measure from a process standpoint. It is guideline based and provides appropriate quantity, quality and consistency of evidence to assess the inappropriate utilization of cardiac imaging fro preoperative risk assessment for non-cardiac, low risk surgery.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***
   - It measures the overuse of cardiac imaging prior to non-cardiac surgery and can be applied directly.

1b. ***Committee's Comments on Performance Gap:***
   - Performance data was provided.
   - Based on the 2013 data, race/ethnicity and facility location (urban vs. rural) had a statistically significant likelihood of undergoing inappropriate imaging therefore indicating a gap in care. Subgroup data was provided indicating a disparity in care.

1c. ***Committee's Comments on Composite Performance Measure:***
   - All performance measures were stated and logical.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The data source includes administrative claims. HCPCS, ICD-9, ICD-10, and CPT codes provided for the numerator denominator and exclusions, with "look-back" timing as applicable. An ICD-10 conversion methodology is not provided. A clear calculation algorithm is provided.
- Since initial endorsement in 2011, 48 new codes added to the list of low-risk surgical procedures included in the numerator.
- Based on a 2013 update to the ACC/AHA guidelines on Perioperative Cardiovascular Evaluation and Care for Non-Cardiac Surgery an exclusion from the measure's denominator was added in 2014 for those patients with 3 of 5 clinical risk factors (i.e., diabetes mellitus, renal insufficiency, stroke or transient ischemic attack, prior heart failure, or ischemic heart disease) to determine non-low risk patients.
- The measure is not risk adjusted, though the developers state the testing sample includes unique patient and facility characteristic coding.
- For this measure, better quality equates to lower scores.

Questions for the Committee:

○ *Are all the data elements clearly defined? Are all appropriate codes included?*

○ *Is the logic or calculation algorithm clear?*

○ *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- Reliability testing for this measure was conducted at the level of the performance measure score, using the data source and level of analysis specified. The primary analysis was conducted at the facility level and included 2013 Medicare FFS data from 2,759 facilities.
- Reliability was calculated using *two* tests: the ability to identify statistical outliers and a signal-to-noise analysis.
  - The statistical outliers – As previously mentioned, one of the intentions of this measure is to identify statistical outliers. Of the 2,759 facilities having met an undetermined minimum count of studies, 137 were classified as outliers (defined as falling outside of the confidence interval (± 1.96 standard deviations) for the measure mean or benchmark value, in 2013; these facilities were reported as having statistically significant rates of overuse.
  - The signal-to-noise analysis, which is appropriate for the measure, quantifies the amount of variation in performance that is due to differences between different facilities (as opposed to differences that are due to random measurement error). The method results in a reliability statistic for each facility. The developers state that the beta-binomial model determined a mean reliability score of 43.0%, which the developer states is indicative of moderate reliability.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The clinical practice guidelines supporting this measure recommend that patients undergoing low-risk, non-cardiac surgery should not have stress image testing.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- Face validity of the measure score and data element was systematically assessed through a 7 member Technical Expert Panel (TEP), majority who were not involved in the development of the measure.
- Overall, the TEP members support the validity of the measure, as specified. 75% of the TEP members agreed the 30-day window used to look forward for a low-risk non-cardiac surgery from the date of the imaging procedure accurately captures preoperative testing. 86% and 71% agreed that the imaging procedures and exclusions, respectively, are accurately captured using claims data.
- The developers state that the TEP was not able to reach a consensus on which clinical conditions should be excluded from the measure. The TEP members felt they did not have the clinical knowledge to provide a definitive response. However, these exclusions are based on the revised AHA/ACC clinical guideline.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The measure excludes patients with three or more of the listed risk factors in the prior three years (diabetes, renal insufficiency, stroke or transient ischemic attack, prior heart failure or ischemic heart disease) as determinants of non-low risk patients.
- The developers tested the prevalence of each exclusion by facility and at an aggregate level; they also tested the aggregate risk factor exclusion to determine the effect on facility performance scores using 2013 data.
- A total of 2,770 facilities in 2013 met an undetermined minimum case count which included a patient population of 647,957 and 633,195 cardiac imaging studies. Overall, there were 395,577 (37.9%) occurrences of any three of the risk factors.
- The developer also calculated descriptive statistics for the measure scores of each facility, with and without, the exclusion for patients with three or more risk factors.
- The frequency of the exclusion for patients with three or more risk factors varied substantially across facilities (IQR: 13.7%). Despite the variation across facilities, the developers state exclusion of patients with 3 or more risk factors is appropriate.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This process measure is not risk adjusted.
- Despite statistically significant differences in imaging prior to low-risk, non-cardiac surgery based on race and facility, the developer believes these differences are driven by a variation in provider practice, therefore, risk adjustment/stratification is not necessary or appropriate.

Questions for the Committee:

- *Does the SC feel the measure is appropriate for risk-adjustment?*

2b5. Meaningful difference:

- The developers state that many facilities had a performance score between 4% and 5% in 2013 and more than 7% of facilities continued to have an outlying performance.
- The developers state that by reporting a measure mean (benchmark value), outlying facilities have the opportunity to identify their high rate of overuse and implement quality improvement strategies to reduce the rate of overuse of imaging studies.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- The developer states that the analytic files used for measure testing and measure calculation include post-adjudicated claims, and do not include missing data.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- Data elements are provided , codes are provided. The 5 clinical risk factors provided although covered in the ACC/AHA guidelines do not comment on patient functional status which may be the most important preoperative assessment.

2a2.: ***Committee's Comments on Reliability-Testing:***

- Yes, reliability testing was performed looking at both statistical outliers and signal-to-noise analysis. The results demonstrate moderate reliability with a score of 43%.

2b1.: ***Committee's Comments on Validity-Specifications***:

- They are not inconsistent with the evidence.
- None

2b2.: ***Committee's Comments on Validity-Testing:***

- Yes validity testing was performed with adequate scope. A 7 member TEP was used and found that 75% agreed on the 30 day window would adequate capture preoperative testing. 86% and 71% agreed that imaging procedures and exclusions were accurately captured.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- The measure excludes patients with 3 or more of the listed risk factors diagnosed in the prior 3 years as determinants of non-low risk patients.
- 2770 facilities met an undetermined minimum case count and overall 395,577 (37%) had occurrences of any three.
- Process measure is not risk adjusted
- Comparibility is not applicable

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- The data source includes administrative claims and uses CMS hospital outpatient claims as its data source; all data elements are in defined fields in electronic claims.
- This is not an eMeasure.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- The data source would include administrative claims using CMS hospital outpatient claims as its data source, so thus feasible.

Criterion 4: Usability and Use

4. Usability and 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.

- The measure is currently publicly reported in CMS' Hospital Outpatient Quality Reporting program.
- The developers state that they did not identify any untended consequences during measure testing.
- NQF's Measure Application Partnership (MAP) reviewed the measure with the following recommendations:

MAP Report Year	Measure Set	Care Setting	Level of Analysis	MAP Finding
2012	Acute Cardiovascular Conditions	Urgent Care	Facility, National	None Provided

***Questions for the Committee**:*

- *Is the measure publicly reported?*
- *For maintenance measures – is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- The measure is currently reported in the CMS HOQR
- No unintended consequences.
- It can reduce unnecessary stress testing.

Criterion 5: Related and Competing Measures

- 0670 : Cardiac stress imaging not meeting appropriate use criteria: Preoperative evaluation in low risk surgery patients
- The developer states that although NQF #0669 is similar to NQF #0670, there are several differences that would make measure harmonization infeasible and reduce the effectiveness of both currently endorsed measures.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0669
Measure Title: Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac, Low Risk Surgery
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: <Date>

Instructions

- *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.

- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Overuse of cardiac imaging prior to non-cardiac, low-risk surgery for patients at low or moderate risk of cardiac involvement.

☐ Structure: Click here to name the structure

☐ Other:

_____

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

This measure is not a health outcome/PRO performance measure.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

This measure is not a health outcome/PRO performance measure.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

The process of identifying low-risk patients undergoing surgeries for which there is low or moderate risk of cardiac involvement prior to performing preoperative cardiac imaging is related to improved outcomes, including reduced exposure to radiation, reduced exposure to contrast agents, and more efficient use of imaging resources.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Two clinical practice guidelines are provided based on their relevance and sample size for the measure. The first guideline, developed by the European Society of Cardiology and the European Society of Anaesthesiology, evaluates cardiovascular assessment and management for non-cardiac surgery. The second guideline, from the American College of Cardiology and American Heart Association, evaluates preoperative cardiovascular evaluation and management of the adult patient undergoing non-cardiac surgery. Citations for the two guidelines follow:

1 – Kristensen SD, Knuuti J, Saraste A, et al. ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management: The Joint Task Force on non-cardiac surgery: cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesthesiology (ESA). European Heart Journal. 2014. Guideline available at: http://eurheartj.oxfordjournals.org/content/ehj/35/35/2383.full.pdf.

2 – ACC/AHA 2014 Guidelines on Perioperative Cardiovascular Evaluation and Care for Non- cardiac Surgery: a report of the American College of Cardiology/American Heart Association task force on practice guidelines. J Am Coll Cardiol. 2007; 50: e159-e242. Guideline available at: http://content.onlinejacc.org/article.aspx?articleid=1893784.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Guideline 1 recommends that patients undergoing low-risk, non-cardiac surgery should not have stress image testing. For intermediate- to high-risk, non-cardiac surgeries, only patients with one to two clinical risk factors and poor functional capacity should undergo imaging stress testing prior to the non-cardiac surgery. *Guideline 1* provides four recommendations to support the measure's clinical intent.

Guideline 2 recommends that patients undergoing low-risk, non-cardiac surgery should not have stress image testing. For patients with elevated risk and poor/unknown functional capacity, it may be reasonable to perform exercise stress-image testing. *Guideline 2* provides five recommendations to support the measure's clinical intent.

Guideline 1 Recommendations:

A. Selected patients with cardiac disease undergoing low-and intermediate-risk non-cardiac surgery may be referred by the anaesthesiologist for cardiological evaluation and medical optimization (Class IIb, Level C; pg. 2388).
B. Imaging stress testing is recommended before high-risk surgery in patients with more than two clinical risk factors and poor functional capacity (<4 METs) (Class I, Level C; pg. 2394).
C. Imaging stress testing may be considered before high- or intermediate-risk surgery in patients with one or two clinical risk factors and poor functional capacity (<4 METs) (Class IIb, Level C; pg. 2394).
D. Imaging stress testing is not recommended before low-risk surgery, regardless of the patient's clinical risk (Class III, Level C; pg. 2394).

Guideline 1: European Society of Cardiology/European Society of Anaesthesiology

Surgical risk estimate according to type of surgery or intervention:

Low Risk (<1%)

- Superficial surgery
- Breast
- Dental
- Endocrine: thyroid
- Eye
- Reconstructive
- Carotid asymptomatic (CEA or CAS)
- Gynaecology: minor
- Orthopaedic: minor (meniscectomy)
- Urological: minor (transurethral resection of the prostate)

Intermediate Risk (1% to 5%)

- Intraperitoneal: splenectomy, hiatal hernia repair, cholecystectomy
- Carotid symptomatic (CEA or CAS)
- Peripheral arterial angioplasty
- Endovascular aneurysm repair
- Head and neck surgery
- Neurological or orthopaedic: major (hip and spine surgery)
- Urological or gynaecological: major
- Renal transplant
- Intra-thoracic: non-major

High-risk (>5%)

- Aortic and major vascular surgery
- Open lower limb revascularization or amputation or thromboembolectomy
- Duodeno-pancreatic surgery
- Liver resection, bile duct surgery
- Oesophagectomy
- Repair of perforated bowel
- Adrenal resection
- Total cystectomy
- Pneumonectomy
- Pulmonary or liver transplant

Guideline #2 Recommendations: American College of Cardiology/American Heart Association

A. For patients with elevated risk and excellent functional capacity, it is reasonable to forgo further exercise testing and proceed to surgery (Class IIa, Level B; pg. e96).
B. For patients with elevated risk and unknown functional capacity it may be reasonable to perform exercise testing to assess for functional capacity if it will change management (Class IIb, Level B; pg. 396).
C. For patients with elevated risk and moderate to good functional capacity, it may be reasonable to forgo further exercise testing and proceed to surgery (Class IIb, Level B; pg. e96).
D. For patients with elevated risk and poor or unknown functional capacity it may be reasonable to perform exercise testing with cardiac imaging to assess for myocardial ischemia (Class IIb, Level C; pg. e96).
E. Routine screening with noninvasive stress testing is not useful for low-risk noncardiac surgery (Class III, Level B; pg. e96).

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Recommendations made within *Guideline 1* ranged from Class I through Class III. The evidence supporting these recommendations demonstrates that cardiac imaging is appropriate for patients with a history of cardiac disease or for whom cardiac involvement is likely, and that it is unnecessary for those patients for whom no risk factors exist.

Recommendations made within *Guideline 2* ranged from Class IIa through Class III. The evidence for these recommendations is supported by either a single randomized control trial or a non-randomized study. Generally, the evidence available suggests consensus within the clinical community that cardiac imaging in patients with low likelihood of cardiac involvement is well accepted (though there may be some conflicting evidence within single randomized or non-randomized studies).

The following grading scale applies to recommendations from *Guideline 1*:
Recommendation A: Class IIb: Usefulness/efficacy is less well established by evidence/opinion.
Recommendation B: Class I: Evidence and/or general agreement that a given treatment or procedure is beneficial, useful, and effective.
Recommendation C: Class IIb: Usefulness/efficacy is less well established by evidence/opinion.
Recommendation D: Class III: Evidence or general agreement that the given treatment or procedure is not useful/effective, and in some cases may be harmful.

The following evidence scales apply to recommendations from *Guideline 1*:
Three levels of evidence: Level A, Level B, and Level C.
Level A: Data derived from multiple randomized clinical trials or meta-analyses.
Level B: Data derived from a single randomized clinical trial or large non-randomized studies.
Level C: Consensus of opinion of the experts and/or small studies, retrospective studies, registries.

Five classes of recommendations: Class I, Class II, Class IIA, Class IIB, and Class III.
Class I: Evidence and/or general agreement that given treatment or procedure is beneficial, useful, and effective.
Class II: Conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of the given treatment or procedure.
Class IIA: Weight of evidence/opinion is in favor of usefulness/efficacy.
Class IIB: Usefulness/efficacy is less well established by evidence/opinion.
Class III: Evidence or general agreement that the given treatment or procedure is not useful/effective, and in some cases may be harmful.

The following grading scale applies to recommendations from *Guideline 2*:
Recommendation A: Class IIa, Level B: Recommendation in favor of treatment or procedure being useful/effective; some conflicting evidence from single randomized trial or nonrandomized studies
Recommendation B: Class IIb, Level B: Recommendation's usefulness/efficacy less well established; Greater conflicting evidence from single randomized trial or nonrandomized studies.

Recommendation C: Class IIb, Level B: Recommendation's usefulness/efficacy less well established; Greater conflicting evidence from single randomized trial or nonrandomized studies.
Recommendation D: Class IIb, Level C: Recommendation's usefulness/efficacy less well established; only diverging expert opinion, case studies, or standard of care.
Recommendation E: Class III, Level B: Recommendation that procedure or treatment is not useful/ineffective and may be harmful; evidence from single randomized trial or nonrandomized studies.

The following evidence scales apply to recommendations from Guideline 2:
Three levels of evidence: Level A, Level B, and Level C.
Level A: Multiple populations evaluated. Data derived from multiple randomized clinical trials or meta-analyses.
Level B: Limited populations evaluated. Data derived from single randomized trial or nonrandomized studies.
Level C: Very limited populations evaluated. Only consensus opinion of experts, case studies or standard of care.

Four classes of recommendations: Class I, Class IIa, Class IIb, Class III.
Class I: Benefit >>> Risk. Procedure/treatment SHOULD be performed/administered.
Class IIa: Benefit >> Risk. Additional studies with focused objectives needed. IT IS REASONABLE to perform procedure/administer treatment.
Class IIb: Benefit ≥ Risk. Additional studies with broad objectives needed; additional registry data would be helpful. Procedure/treatment MAY BE CONSIDERED.
Class III: No benefit or harmful.

- Procedure/Test – Not helpful OR Excess cost with no benefit or harmful
- Treatment – No proven benefit OR harmful to patients

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)
All relevant information about the grades and associated definitions for the two guidelines provided has been included in **section 1a.4.3.**

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**
Citations and URLs are the same as those noted in **section 1a.4.1**.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?
☒ Yes → ***complete section 1a.7***
☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION
1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):
This measure is not based on a United States Preventive Services Task Force Recommendation.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.
This measure is not based on a United States Preventive Services Task Force Recommendation.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:
This measure is not based on a United States Preventive Services Task Force Recommendation.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)
This measure is not based on a United States Preventive Services Task Force Recommendation.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**
This measure is not based on a United States Preventive Services Task Force Recommendation.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE
1a.6.1. Citation (*including date*) and **URL** (*if available online*):
Guidelines are evidenced based; details are provided in **section 1a.7**.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**
Guidelines are evidenced based; details are provided in **section 1a.7**.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE
If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?
The clinical focus of *Guideline #2*, for which this systematic review was performed, is the perioperative cardiovascular evaluation and management of the adult patient undergoing non-cardiac surgery, including preoperative risk assessment and cardiac imaging. Preoperative cardiac imaging can serve three purposes, including 1) assessment of perioperative risk, 2) determination of the need for changes in management, and 3) identification of cardiovascular conditions or risk factors requiring longer-term management. Results of the cardiac imaging may lead to recommendations and discussions with the perioperative team about the optimal location and timing of surgery (e.g., ambulatory surgery center versus outpatient hospital, or inpatient admission) or alternative strategies to approaching patient care.

Methodologic Approach for the Systematic Review that Supports *Guideline #2*:
In April 2013, the American College of Cardiology/American Heart Association task force on practice guidelines conducted an extensive evidence review, which included a literature review for evidence published through July 2013. Other selected references published through May 2014 were also incorporated by the guideline writing committee (GWC). Literature included was derived from research involving human subjects, published in English, and indexed in MEDLINE (through PubMed), EMBASE, the Cochrane Library, Agency for Healthcare Research and Quality Reports, and other selected databases relevant to this clinical practice guideline (CPG). The relevant data are included in evidence tables in the ACC/AHA Data Supplement available online (link here: http://jaccjacc.cardiosource.com/acc_documents/2014_Periop_GL_Data_Supplement_Tables.pdf); seven articles were identified by the ACC/AHA task force that were relevant to *preoperative exercise stress testing for myocardial ischemia and functional capacity*, which are located on Data Supplement 11.

Key search words included, but were not limited to, the following: anesthesia protection; arrhythmia; atrial fibrillation; atrioventricular block; bundle branch block; cardiac ischemia; cardioprotection; cardiovascular implantable electronic device; conduction disturbance; dysrhythmia; electrocardiography; electrocautery; electromagnetic interference; heart disease; heart failure; implantable cardioverter-defibrillator; intraoperative; left ventricular ejection fraction; left ventricular function; myocardial infarction; myocardial protection; National Surgical Quality Improvement Program; pacemaker; perioperative; perioperative pain management; perioperative risk; postoperative; preoperative; preoperative evaluation; surgical procedures; ventricular premature beats; ventricular tachycardia; and volatile anesthetics.

An independent evidence review committee (ERC) was commissioned to perform a systematic review of a key question, the results of which were considered by the GWC for incorporation into this CPG. Please refer to the ACC/AHA task force's systematic review report, published in conjunction with this CPG and its respective data supplements.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:
Grade for the evidence provided from *Guideline #2* can be found in **section 1a.4.3.**

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.
Grade for the evidence provided from *Guideline 2* can be found in section **1a.4.3.**

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Guideline #2: 2007 - 2013

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)
Guideline #2: 213 articles were included in the body of evidence, 7 of which were directly related to preoperative stress testing for noncardiac low-risk surgery. Of these seven, two of the studies were observational, two were prospective, two were consecutive case series, and one was a retrospective analysis.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)
Consistency was seen in the outcomes of the seven studies cited by *Guideline 2* within the ACC/AHA task force Data Supplement. Of the seven studies cited, some limitations were identified, including small sample sizes for each study and potential concerns with patient assignment into the experimental and control arms. Collectively, however, the results cited within the body of evidence for the recommendations made in *Guideline #2* were consistent across studies and were of moderate quality.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)
Given the high costs associated with performing unnecessary cardiac stress testing prior to non-cardiac, low-risk surgeries, the increased level of radiation to which these patients are exposed unnecessarily, and the fact that preoperative stress testing prior to non-cardiac, low-risk surgery is unlikely to change the outcome of the patient's operative or treatment plan. The overall net benefit in reducing overuse of cardiac stress testing prior to non-cardiac, low-risk surgery is a reduction in cost, a reduction in radiation exposure, and a reduction in the number of procedures performed, per beneficiary.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?
No harms in measure implementation were identified to counter the net benefit of the measure.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.
In addition to the two guidelines cited above, a review of the clinical literature was conducted during the measure contractor's annual review of the literature for additional evidence and/or new studies that substantiate the measure's intent. Citations and summaries for the 14 items included in this review can be found in **Section 1a.8.2**. Some of these 14 studies have been published since the period of guideline development. Results cited in these studies are consistent across studies and with the guidelines cited above.

1a.8 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.8.1 What process was used to identify the evidence?

In addition to the two guidelines cited above, a review of the clinical literature was conducted during the measure contractor's annual review of the literature for additional evidence and/or new studies that support the measure's intent. The measure contractor identified relevant peer-reviewed publications by searching the PubMed MEDLINE database from January 1, 2013 to January 16, 2015, limiting included results to those published in the English language and that had abstracts available in PubMed. The search initially identified 96 articles; a further review by the contractor's clinical and measure-development team resulted in the inclusion of 14 articles in the body of evidence below. Citations and summaries for the 14 items included in this review can be found in **Section 1a.8.2**.

1a.8.2. Provide the citation and summary for each piece of evidence.

Augoustides JG, Neuman MD, Al-Ghofaily L, Silvay G. Preoperative cardiac risk assessment for non-cardiac surgery: defining costs and risks. J Cardiothorac Vasc Anesth. 2013; 27(2):395-9.

Augoustides et al. reviewed literature on the appropriateness of preoperative cardiac risk assessment prior to non-cardiac surgery. The review notes that cardiac stress testing in Ontario acute care hospitals was associated with harm in low-risk patients (hazard ratio 1.35; 95% CI 1.05–1.74). However, preoperative cardiac stress testing was associated with benefits for intermediate-risk (hazard ratio 0.92; 95% CI 0.85–0.99) and high-risk patients (hazard ratio 0.80; 95% CI 0.67–0.97). Reducing cardiac stress testing prior to non-cardiac surgery for low-risk cases presents a cost-saving opportunity and a reduction in unnecessary testing. The study also notes that risk stratification tools such as the thoracic risk index effectively stratify patient risks and identify patients for whom additional cardiac stress testing may be appropriate.

Carryer D, Hodge D, Miller T, Askew J, Gibbons R. Application of appropriateness criteria to stress single photon emission computed tomography sestamibi studies: a comparison of the 2009 revised appropriateness criteria to the 2005 original criteria. Am Heart J. 2010; 160:244–9.

Carryer et al. retrospectively examined 281 patients who underwent stress SPECT MPI at the Mayo Clinic between May 1, 2005, and May 15, 2005. Using the revised 2009 ACCF/ASNC AUC, the researchers compared these findings to previously published results in this same cohort using the 2005 AUC. The most common broad indication for testing was post-revascularization (34 percent of the group), followed by evaluation of chest pain or ischemic equivalent (25 percent), follow-up of prior testing (20 percent), and screening of asymptomatic patients (14 percent). Evaluation for preoperative assessment for non-cardiac surgery (6 percent) was uncommon. Compared to the 2005 AUC, the 2009 AUC resulted in a highly significant overall change ($P < .001$) in the classification of appropriateness. The 2009 AUC eliminated unclassified patients, reduced appropriate studies (59.8 percent versus 63.7 percent, $P = .02$), increased studies of uncertain appropriateness (16.0 percent versus 10.7 percent, $P = .01$), and increased inappropriate studies (24.2 percent versus 14.6 percent, $P < .001$).

Colla CH, Morden NE, Sequist TD. Choosing Wisely: prevalence and correlates of low-value health care services in the United States. J Gen Intern Med. 2014; 30(2): 221-228.

Colla et al. aimed to develop claims-based algorithms to estimate the prevalence of select Choosing Wisely services and to examine the relationship between low-value care at the regional level and demographic, health, and healthcare system differences. The study team developed claims-based algorithms to measure the prevalence of 11 low-value services as identified by Choosing Wisely and examined geographic variation across hospital referral regions for these services. The study team found that the national average annual prevalence of the measured Choosing Wisely low-value services ranged from 1.2 percent to 46.5 percent and that prevalence across hospital referral regions varied significantly. For preoperative cardiac testing for non-cardiac surgery, specifically, the study team found that the average annual prevalence was 46.5 percent, with an estimated waste of $3.2 million. The study team stated that potential implications of these findings are that non-indicated

preoperative cardiac testing could lead to patient harm from additional testing or an increase in false positive results. The study team concluded that identifying and measuring low-value health services could help increase the value of healthcare and improve services and areas with greater use of potentially inappropriate care.

Ghadri JR, Fietcher M, Veraguth K, et al. Coronary calcium score as an adjunct to nuclear myocardial perfusion imaging for risk stratification before non-cardiac surgery. J Nucl Med. 2012; 53(7):1081-6.

Ghadri et al. studied the value of coronary artery calcium score (CACS) as an addition to SPECT MPI for cardiac risk stratification prior to non-cardiac surgery. The study included 326 patients referred for SPECT MPI prior to elective non-cardiac surgery, who then also underwent a low-dose CT scan to determine the CACS. CACS was found to be a strong preoperative risk predictor independent of the SPECT MPI study. CACS was also found to add additional risk stratification information to the results of SPECT MPI studies. In particular, patients with a normal SPECT MPI result, were still at increased risk for a major adverse cardiac event if the result of the CACS test was found to be above the normal threshold ($p = 0.003$).

Gibbons RJ, Miller TD, Hodge D, Urban L, Araoz PA, Pellikka P, McCully RB. Application of appropriateness criteria to stress single-photon emission computed tomography sestamibi studies and stress echocardiograms in an academic medical center. J Am Coll Cardiol. 2008; 51(13).

Gibbons et al. retrospectively applied the 2005 ACCF/ASNC AUC to 284 patients who underwent stress SPECT MPI and 298 patients who underwent stress echocardiography in a single academic medical center. Fourteen percent of stress SPECT studies and 18 percent of stress echo studies were performed for inappropriate reasons. Stress SPECT patients were more likely to be referred for follow-up testing post-revascularization and for follow-up testing after prior SPECT testing. In contrast, stress echo patients were more likely to be referred for preoperative assessment before non-cardiac surgery and the assessment of symptoms. Very few patients were referred for risk assessment after acute coronary syndromes. Similar percentages (14–15 percent) of both groups were asymptomatic patients without other indications, who were being screened for suspected coronary artery disease.

Hashimoto J, Nakahara T, Bai J, et al. Preoperative risk stratification with myocardial perfusion imaging in intermediate and low-risk non-cardiac surgery. Circ J. 2007; 71:1395–1400.

Hashimoto et al., in a retrospective cohort study, examined data collected from patients referred for preoperative stress myocardial perfusion SPECT before non-cardiac surgery and concluded that preoperative SPECT provides an incremental prognostic value in intermediate, but not in low-risk surgery. More than 1,200 consecutive patients underwent electrocardiography-gated dipyridamole stress SPECT to evaluate myocardial perfusion and cardiac function before intermediate- or low-risk non-cardiac surgery. Perfusion or cardiac function yielded significant risk stratification in intermediate- but not low-risk surgery. Adding functional data to perfusion variables offered an incremental prognostic value for patients with an intermediate clinical risk and scheduled intermediate risk surgery.

Hendel RC, Cerqueira M, Douglas PS, Caruth KC, Allen JM, Jensen NC, Pan W, Brindis R, Wolk M. A multicenter assessment of the use of single-photon emission computed tomography myocardial perfusion imaging with appropriateness criteria. J Am Coll Cardiol. 2010; 55(2).

Hendel et al. assessed the feasibility of evaluation for appropriate use of SPECT MPI in multiple clinical sites to determine use patterns as well as to identify areas of apparent inappropriate use. Six diverse clinical sites enrolled consecutive patients undergoing MPI. An automated algorithm assigned a specific indication from the 2005 ACCF/ASNC appropriateness criteria for SPECT MPI. Of the 6,351 patients enrolled in the study, 93 percent were successfully assigned an appropriateness level. Inappropriate use of MPI was found in 14.4 percent of patients, with a range of 4–22 percent among practices. Women and patients less than 65 years of age were more likely to undergo inappropriate MPI. Asymptomatic, low-risk patients accounted for 44.5 percent of inappropriate testing.

Henzlova M, Savino J, Levine EJ, Croft L, Einstein A, Hermann L, Duvall W. Comparative effectiveness of coronary CT angiography versus stress testing using high-efficiency SPECT myocardial perfusion imaging and stress-only imaging in the emergency department. J Am Coll Cardiol. 2013; 61(10_S).

Henzlova et al. compared CT angiography (CTA) and stress testing with SPECT MPI. During a period of two years, 1,458 individuals underwent testing in the emergency department (ED). A total of 192 CTAs and 1,266 stress tests (939 MPIs and 327 exercise treadmill testing (ETTs) were performed. In general, patients undergoing CTA were a lower risk group, based on age, risk factors, and known presence of heart disease. Within the MPI cohort, 708 underwent stress-only imaging. The percentage of patients discharged directly from the ED by stress-testing group, as compared with CTA group, was 82 percent versus 73 percent. However, the time to disposition was significantly longer for the stress-testing group as compared with the CTA group (20.5 ± 7 versus 11.0 ± 5 hrs, $p < 0.0001$). Additionally, there were more cardiac return visits to the ED in individuals who underwent CTA as opposed to stress testing (47 percent versus 10 percent, $p = 0.0002$). In general, the study concluded that stress testing with ETT, high-efficiency SPECT MPI, had significantly lower patient radiation dose, fewer cardiac return visits, and fewer follow-up diagnostic testing than CTA. Additionally, while CTA had a lower time to disposition than stress testing, there was a trend toward more discharges from the ED with stress testing.

Kerr EA, Chen J, Sussman JB, et al. Stress testing before low-risk surgery: so many recommendations, so little overuse. JAMA Internal Medicine. 2015; E1-E2.

Kerr et al. sought to determine the prevalence of cardiac stress testing before low-risk surgeries, prior to commencement of the Choosing Wisely campaign, to estimate the effect of the recommendations on future use of resources. The study team performed a retrospective cohort study using data from the Department of Veterans Affairs (VA) and from a nationally representative five percent sample of Medicare fee-for-service claims. The study team found that a routine preoperative stress test preceded 1 of the 3 low-risk surgeries in only 0.67 percent of VA patients and 2.14 percent of Medicare patients. The study team concluded that routine preoperative stress testing before low-risk surgery in both VA and Medicare patients is low and shows little variance across geographic regions. These trends existed even prior to the Choosing Wisely campaign, so the study team postulated that most physicians had already been incorporating guidelines about appropriate preoperative stress testing into their practices before the Choosing Wisely recommendations were developed.

Koh AS, Flores JL, Keng FY, Tan RS, Chua TS. Correlation between clinical outcomes and appropriateness grading for referral to myocardial perfusion imaging for preoperative evaluation prior to non-cardiac surgery. J Nucl Cardiol. 2012; 19(2):277–84.

Koh et al. studied the correlation of appropriateness of grading with both the outcome of MPI as well as the clinical outcomes of patients post-surgery. Patients visiting the MPI laboratory from March 2009 to July 2009 were studied and, based on their medical records, stress data, and imaging results, MPI scans were classified as appropriate, inappropriate, uncertain, or unclassified. Based on AUC, MPI referrals in intermediate- and high-risk groups with poor functional class were graded as appropriate, while MPI referrals in low-risk and intermediate risk groups with normal functional class were graded as inappropriate referrals. Out of 176 referrals for preoperative evaluation, 39.8 percent were graded as inappropriate. The study concluded that MPI results predicted outcome in appropriately tested patients, but not in patients whose tests were classified as inappropriate.

Koh AS, Flores JL, Keng FY, Tan RS, Chua TS. Evaluation of the American College of Cardiology Foundation/American Society of Nuclear Cardiology appropriateness criteria for SPECT myocardial perfusion imaging in an Asian tertiary cardiac center. J Nucl Cardiol. 2011 Apr; 18(2):324–30.

Koh et al. studied the pattern of referrals for SPECT MPI in an Asian tertiary cardiac center. Medical records and stress data were studied for 1,623 consecutive patients (mean age 61 years, 61 percent males) who were referred to the MPI laboratory between February 16 and June 19, 2009. MPI studies were categorized into appropriate, inappropriate, uncertain, or unclassified according to the 2009 ACCF/ASNC appropriateness criteria for SPECT MPI. Most common indications for SPECT were evaluation of ischemic equivalent for coronary artery disease, risk assessment post-revascularization, and preoperative evaluation for non-cardiac surgery. Ten percent of referrals were classified as inappropriate, 5 percent as uncertain, and 3 percent as unclassified. The

preoperative group had the highest proportion of inappropriate studies (59 percent). The authors concluded that preoperative evaluation for low-risk surgery appeared to be the most common source of inappropriate referrals for SPECT MPI in their institution.

McCully RB, Pellikka PA, Hodge DO, Araoz PA, Miller TD, Gibbons RJ. Applicability of appropriateness criteria for stress imaging: similarities and differences between stress echocardiography and single-photon emission computed tomography myocardial perfusion imaging criteria. Circ Cardiovasc Imaging. 2009; 2:213–8.

McCully et al. evaluated the application of the stress echocardiography appropriateness criteria to patients undergoing stress echocardiography in an academic medical center. The stress echocardiography criteria were applied to 298 consecutive patients who underwent stress echocardiography. Results were compared with those of a previous analysis in the same patients using the SPECT MPI criteria. Overall, 54 percent of patients were classified as appropriate, 8 percent as uncertain, and 19 percent as inappropriate; 19 percent were not classifiable. By the SPECT MPI criteria, 64 percent of patients were classified as appropriate, 9 percent as uncertain, and 18 percent as inappropriate; 9 percent were not classifiable ($P < .001$ compared with stress echocardiography criteria). By the stress echocardiography criteria, preoperative evaluation for low-risk surgery in patients with minor or intermediate clinical risk predictors was among six clinical situations or indications that accounted for more than 90 percent of the inappropriate tests. Most of these involved asymptomatic patients.

Placanica G, Merola R, Placanica A, Pecoraro A, Fusco L, Placanica P, Pasta V. Cardiological assessment of cardiac patients undergoing non-cardiac surgery (usefulness of surveys). Ann Ital Chir. 2011 May–Jun; 82(3):179–84.

Placanica et al. studied the effective usefulness of preoperative stress test and echocardiography in adult patients with coronary artery disease, undergoing non-cardiac surgery. Two-hundred patients age 58–85, affected by stable ischemic pathology, undergoing non-cardiac surgery, and treated with oral drugs, were enrolled for an assessment protocol including a preliminary case history (anamnesis); objective examination; blood pressure; race, class, and gender blood chemistry analysis; and cardiac risk evaluation. A second cohort of 50 patients with similar characteristics was subjected to the same tests and preoperative and exercise stress test. All patients showed a good hemodynamic compensation and a quick recovery. The group of 200 patients, for whom the risk assessment was performed without preoperative and stress test, concluded the process three days prior, on average, before undergoing an echocardiography and exercise stress test. The authors concluded that when patients are hemodynamically stable and their conditions controlled by appropriate therapy, it is sufficient to perform first-level tests for the preoperative stratification of cardiovascular risk. It is recommended to perform echocardiogram and stress test when the first level tests are abnormal, when there is a worsening of the conditions prior to admission, or when the patient is not hemodynamically stable.

Sumin AN, Korok EV, Kokov AN, et al. Role of multidetector computed tomography and stress- echocardiography in assessment of risk of cardiological complications of non-cardiac surgery. Kardiologiia. 2014; 54(5): 39-47.

Sumin et al. sought to compare results of coronary angiography with data of multi-slice computed tomography (MSCT) angiography, and analyzed the rate of detection of hemodynamically significant coronary artery lesions during preoperative examination of patients. The investigators analyzed case histories of 92 patients [median age 59 years] examined prior to surgery on non-coronary vessels, or for exclusion of ischemic heart disease. All patients were subjected to selective coronary angiography (CA) and MSCT angiography. According to results of CA patients were divided into two groups – those with coronary artery stenoses >70% (n=55, group 1) and <70% (n=37, group 2). In 46 patients (50%), dobutamine stress echocardiography was performed for detection of concealed coronary insufficiency. The study concluded that the diagnostic value of MSCT for detection of coronary artery stenoses >70% have better sensitivity, specificity, and negative and positive predictive value compared with stress echocardiography. The authors concluded that the study confirms the value of MSCT angiography for diagnosis of coronary atherosclerosis preoperatively, and recommended it as a screening method for detecting hemodynamically significant coronary artery involvement before extracardiac surgery.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
NQF_0669_Measure_Evidence_Form_2015-06-30.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
This measure aims to reduce overuse of cardiac imaging prior to low-risk non-cardiac surgeries, for patients with low or moderate cardiac risk, as cardiac imaging in this population can result in increased exposure to radiation with little or no clinical benefit. The measure score will guide patient selection of providers, assess quality, and inform quality improvement.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Analysis of Medicare fee-for-service (FFS) claims data indicates variation in the use of cardiac imaging prior to non-cardiac, low-risk surgery. For the time period (July 2013 through June 2014), performance rates ranged from 0.00 percent to 18.0 percent, with a weighted mean of 5.07 percent.

The data presented below represent information for the 1,953 facilities whose denominator counts met minimum case count requirements for all years included in the table.

Further details on the descriptive statistics for longitudinal facility performance are included below:

	2009*	2010*	2011*	2012-2013**	2013-2014**	% Change (2009-2014)
Facilities	1,953	1,953	1,953	1,953	1,953	-
Minimum Value	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1st Percentile	0.9%	1.2%	1.4%	1.2%	0.0%	-0.9%
5th Percentile	2.2%	2.4%	2.7%	2.5%	1.7%	-0.5%
10th Percentile	2.8%	3.1%	3.3%	3.1%	2.5%	-0.3%
25th Percentile	3.8%	4.2%	4.3%	4.1%	3.7%	-0.1%
Median	4.9%	5.5%	5.4%	5.2%	4.8%	-0.1%
75th Percentile	6.3%	6.7%	6.7%	6.4%	6.1%	-0.2%
90th Percentile	7.7%	8.1%	8.0%	7.7%	7.5%	-0.2%
95th Percentile	8.7%	9.2%	8.9%	8.5%	8.5%	-0.2%
99th Percentile	10.5%	11.6%	11.6%	10.7%	11.1%	0.6%
Maximum Value	14.5%	16.0%	17.0%	15.1%	18.0%	3.5%
Mean Performance (Standard Deviation)	5.1% (2.0)	5.6% (2.1)	5.6% (2.1)	5.3% (1.9)	5.0% (2.1)	-0.1% (0.1)
Number of Imaging Studies (Denom)	622,774	676,847	791,191	845,928	556,825	-65,949

*The measurement period for HOQR data reported from 2009 through 2011 ran from January to December.
**Beginning in 2012, the measurement period for HOQR was adjusted to run from July to June; consequently, data are not reported for January through June 2012.

One of the intentions for reporting this measure is to identify facilities with significant outlying performance. As shown in the table above, many facilities cluster around a value of 4 to 5 percent; however, outlying performance persists, indicating there are facilities for which there is a notable rate of overuse.

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.
Data have been included in Section 1b.2; these data represent national performance over time, from 2009 to 2014.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Using 2013 performance data, we evaluated the effect of patient and facility characteristics on the likelihood of each beneficiary having a cardiac imaging procedure occur prior to a non-cardiac, low-risk surgery. Using a logistic regression model, we assessed the impact of patient and facility characteristics for the 662,905 cardiac-imaging procedures performed in 2013 and found that race/ethnicity and facility characteristics had a significant relationship with the rate of inappropriate preoperative cardiac imaging.

The primary finding from the regression model is that patient race/ethnicity has a statistically significant effect on the likelihood that a patient undergoes inappropriate preoperative imaging. African-Americans were slightly less likely to undergo inappropriate preoperative cardiac imaging compared to White beneficiaries (OR 0.948, p=0.021); conversely, Asian beneficiaries were slightly more likely to undergo inappropriate preoperative cardiac imaging prior to a low-risk surgery (OR 1.112, p=0.023).

Facility characteristics also played a role in determining whether a non-cardiac, low-risk surgery was performed following a cardiac-imaging study. When compared to facilities with fewer than 50 beds (a proxy for facility size), facilities with 101 to 250 beds (OR 1.151, p=0.018) and 500 or more beds (OR 1.148, p=0.029) were more likely to perform pre-operative imaging. Similarly, a facility's urbanicity impacted a beneficiary's likelihood of having cardiac imaging performed preoperatively—urban facilities were more likely than rural facilities to be associated with pre-surgical imaging (OR 1.067, p=0.006).

While the regression model identified subpopulations of patients and facilities for which there are statistically significant differences in imaging prior to low-risk, non-cardiac surgery, these disparities do not indicate a need for adjustment of the measure specifications. Adjusting for these differences would mask underlying differences in quality of care. As this is a process measure, there should be no difference in the standard of care for these patients; we believe these statistically significant differences are driven by variation in provider practice. Consequently, we do not believe risk adjustment or stratification is necessary or appropriate for this measure.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
A review of the literature for population-based disparities in the inappropriate use of cardiac imaging prior to low-risk, non-cardiac surgery found one study on the topic. Sheffield et al. (2013) analyzed the overuse of preoperative stress testing and found that inappropriate stress testing prior to low-risk surgery was more likely for females. The study also found that living in a region with higher Medicare expenditures was associated with higher rates of preoperative stress testing.

There is additional literature on population disparities for the use of diagnostic imaging more broadly. Hendel et al (2010) assessed the use of single photon emission computed tomography myocardial perfusion imaging (SPECT MPI), finding that women and younger patients were more likely to undergo inappropriate SPECT MPI studies. This study, however, is not restricted to the Medicare population, and examines inappropriate SPECT MPI use for a variety of cases, not just prior to low-risk surgery.

Lucas et al. (2006) evaluated the utilization of imaging stress tests in the Medicare population, finding that non-black males have the highest rate of utilization, followed by non-black females, black males, and finally black females. The study indicates a disparity in utilization based on both race and gender; one limitation of this study is that it does not report outcomes based on income, age, or other factors. While the study demonstrates a gap in utilization of diagnostic imaging, it does not indicate whether the utilization is considered appropriate or inappropriate.

Sistrom et al. (2012) examined diagnostic imaging utilization, finding that race was a determinant in the likelihood of receiving a diagnostic imaging test. While these disparities were noted within the literature, a review of the measure's specifications did not identify conceptual or empirical evidence suggesting that adjustment based on patient racial or ethnic identification would be indicated for the calculation of provider performance. Thus, this process measure does not require additional adjustment based on sociodemographic factors.

REFERENCES
1.) Hendel R, Cergueira M, Douglas P, et al. A multicenter assessment of the use of single-photon emission computed tomography myocardial perfusion imaging with appropriateness criteria. J Am Coll Cardiol. 2010; 55(2):156-62.

2.) Lucas L, DeLorenzo M, Siewers A, Wennberg D. Treatments for Cardiovascular Disease in the United States, 1993-2001. Circulation. 2006; 113:374-379.
3.) Sheffield K, McAdams P, Benarroch-Gampel J, et al. Overuse of preoperative cardiac stress testing in Medicare patients undergoing elective noncardiac surgery. Ann Surg. 2013; 257(1):73-80.
4.) Sistrom C, McKeay N, Weilburg J, et al. Determinants of diagnostic imaging utilization in primary care. Am J Manag Care. 2012; 18(4): e135-44.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, Frequently performed procedure, High resource use, Patient/societal consequences of poor quality, Other

1c.2. If Other: Safety

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Cardiac imaging is among the most common imaging services in the Medicare population; its use has grown significantly in the past decade. In 2008, hospital outpatient settings performed over 760,000 stress echocardiography, SPECT MPI, and stress MRI procedures. Between 1998 and 2006, the rate of MPI use in Medicare beneficiaries increased 51 percent among cardiologists in the hospital setting, and 215 percent in private offices (Levin 2009). During this period, total Medicare Part B payments for MPI across all settings of care increased by 227 percent (The Lewin Group 2009).

Concomitant with the growth in cardiac imaging, the number of non-cardiac, low-risk surgeries and procedures has progressively increased over the past twenty years (Hernandez 2004). Elderly patients undergo at least four million major, non-cardiac operations annually (Gregoratos 2008). Increased utilization of imaging services also poses a safety concern for patients. Radiation exposure from medical imaging procedures has increased rapidly in recent years, though the growth was fairly consistent across sociodemographic groups from 1997 to 2007. Annual radiation exposure per radiation inducing imaging service procedure increased by 164 percent in emergency departments and 90 percent in physician offices from 1997 to 2007 (Brenner 2007).

While it is important to perform a cardiac risk assessment prior to surgery to identify high-risk patients, in most patients, an extensive cardiac workup is unnecessary, costly, and delays definitive patient care. Perioperative risk is proportional both to the severity of the patient's heart failure and the surgical risk (Savino 2007).

A study evaluating measure concepts included in the Choosing Wisely campaign demonstrated that there is national overuse of cardiac imaging prior to low-risk surgery. Colla et al. (2014) found that the national average annual prevalence of the measured Choosing Wisely low-value services ranged from 1.2 percent to 46.5 percent and that prevalence across hospital referral regions varied significantly. In terms of preoperative cardiac testing for non-cardiac surgery specifically, the study team found that the average annual prevalence was 46.5 percent, with an estimated waste of $3.2 million. Potential implications of these findings suggest non-indicated, preoperative cardiac testing could lead to patient harm from additional testing or an increase in false-positive results. The study team concluded that identifying and measuring low-value health services could help increase the value of health care, improving services and areas with greater use of potentially inappropriate care (Colla et al. 2014).

In general, as evidenced in clinical guidelines and the peer-reviewed literature, preoperative cardiac tests should be performed only if their results are likely to influence patient treatment. Cardiac intervention is rarely necessary to reduce the risk of surgery. The Cleveland Clinic states that, "there are very few cases in which the surgical outcomes and treatments are affected by extensive preoperative cardiac testing. Although preoperative testing is indicated in some cases, it does not always lead to a scientifically measurable improvement in outcome. Indiscriminate and extensive preoperative cardiac testing is an ineffective way of using health care funds and can lead to more unwarranted and risky procedures. In addition to the inappropriate expenditure of resources, unnecessary testing could cause harm to the patient by delaying surgery. For a test to be considered useful, it should be accurate, influence patient outcomes, and have a favorable risk-to-benefit ratio. Therefore, it is essential for the physician to identify patients who will benefit most from an in-depth preoperative cardiac evaluation. It is important for the physician to explore noncardiac issues (e.g., chronic lung disease, coagulopathy, anemia, renal and cerebrovascular disease, diabetes) that can negatively affect the

outcome of the surgery" (Grasso and Wael 2014).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1.) Brenner D, Hall E. Computed tomography - an increasing source of radiation exposure. N Engl J Med. 2007; 357:2277-84.
2.) Colla CH, Morden NE, Sequist TD. Choosing Wisely: prevalence and correlates of low-value health care services in the United States. J Gen Intern Med. 2014; 30(2): 221-228.
3.) Grasso A and Wael J. Cleveland Clinic Center for Continuing Education: Cardiac Risk Stratification for Noncardiac Surgery. 2014.
4.) Gregoratos G. Current guideline-based preoperative evaluation provides the best management of patients undergoing non-cardiac surgery. Circulation 2008; 117(24): 3134-44. Citing: National Center for Health Statistics. Health, United States 2006: inpatient surgery. November 2006.
5.) Hernandez AF, Newby KI, O'Connor CM. Preoperative evaluation for major non-cardiac surgery. Arch Intern Med. 2004; 164: 1729 –1736.
6.) Levin D, Rao V, Parker L, et al. Recent payment and utilization trends in radionuclide myocardial perfusion imaging: Comparison between self-referral and referral to radiologists. J Am Coll Radiol. 2009; 6:437-41.
7.) Savino J and Fleisher LA. Assessment of patients with heart disease for fitness for non-cardiac surgery. Essential Cardiology: Principles and Practice. Ed. Rosendorf. 2nd ed. 2007.
8.) The Lewin Group. NQF Supplemental Preoperative Cardiac Imaging for Low-Risk Surgery, analysis of Medicare Calendar Year 2007 claims data prepared for the Centers for Medicare & Medicaid Services. 2009. HHS Contract No: HHSM-500-2005-0024I, Order No. 0002.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

This measure is not a PRO-PM measure.

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 subcriteria 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 : Screening, Surgery, Surgery : General Surgery

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

Overuse, Safety

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

https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228695266120

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** NQF_0669_Measure_Value_Sets_2015-06-30.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.

As part of the annual measure maintenance and review process, several new codes for low-risk surgeries were added to the numerator in 2012. The 48 new codes added to the list of surgical procedures included in the numerator (see Other Surgeries sub-heading) align the procedure-code list with NQF measure #0670 (cardiac stress imaging not meeting appropriate use criteria: preoperative evaluation in low risk surgery patients). All 48 new surgical procedures fall into the low risk surgical category (i.e., each procedure has a less than one percent surgical-risk estimate). When presented to the contractor's Technical Expert Panel (TEP), the TEP supported the addition of these codes to the numerator, based on a desire to harmonize the list of procedure codes with NQF #0670 and the surgery-risk estimate for each procedure.

Similarly, exclusion from the measure's denominator for those patients identified as at high risk for cardiac involvement was added to the specifications in 2014, based on evidence from the literature and review by the contractor's TEP. The 2013 update to the Guidelines on Perioperative Cardiovascular Evaluation and Care for Non-Cardiac Surgery, developed by the American College of Cardiology/American Heart Association task force on practice guidelines, indicated that patients presenting with three of five concomitant clinical risk factors (i.e., diabetes mellitus, renal insufficiency, stroke or transient ischemic attack, prior heart failure, or ischemic heart disease) were at increased cardiac risk during surgery; consequently, pre-operative cardiac imaging for these patients may be appropriate. Based on this updated evidence, the contractor's TEP recommended adding patients with at least three of these five risk factors to the list of denominator exclusions.

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

The number of stress echocardiography, SPECT MPI, and stress MR studies performed in a hospital outpatient department within 30 days of an ambulatory non-cardiac, low-risk surgery performed at any location (e.g., same hospital, other hospital, or physician office).

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*

Numerator: Stress echocardiography, SPECT MPI, or stress MR procedures, occurring within 30 days of an ambulatory non-cardiac, low-risk surgery, within a 12-month time window (July 1 – June 30).
Denominator: Stress echocardiography, SPECT MPI, or stress MR procedures within an 11-month time window (July 1 – May 31).

S.6. **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, 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.

The numerator is defined by the following categories of surgical procedures:

-Surgery/Integumentary System: Breast
-Surgery/Respiratory System: Accessory Sinuses
-Surgery/Respiratory System: Larynx
-Surgery/Respiratory System: Trachea and Bronchi
-Surgery/Respiratory System: Lungs and Pleura
-Surgery/Digestive System: Esophagus
-Surgery/Digestive System: Intestines (Except Rectum)
-Surgery/Digestive System: Rectum
-Surgery/Digestive System: Anus
-Surgery/Digestive System: Biliary Tract
-Surgery/Digestive System: Abdomen, Peritoneum, and Omentum
-Surgery/Urinary System: Kidney
-Surgery/Urinary System: Ureter
-Surgery/Urinary System: Bladder
-Surgery/Female Genital System: Cervix Uteri
-Surgery/Female Genital System: Corpus Uteri

-Surgery/Female Genital System: Oviduct/Ovary
-Surgery/Eye and Ocular Adnexa: Anterior Segment
-Other Surgeries

(Specific CPT codes for each condition class are included in the value set for this measure; this detailed list can be found in the Excel workbook provided for Section S2b.)

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The number of stress echocardiography, SPECT MPI, and stress MR studies performed in a hospital outpatient department on Medicare beneficiaries within a 12-month time window.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

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

SPECT MPI
CPT 78464, 78451, 78465, 78452

Stress Echocardiography
CPT 93350 C8928 and 93351 C8930

Stress MR
CPT 75559, 75560, 75563, 75564

Global and technical-component (TC) claims should be considered to capture all outpatient volume facility claims, typically paid under the Outpatient Prospective Payment System(OPPS)/Ambulatory Payment Classifications (APC) methodology, and to avoid double counting of professional-component claims (i.e., 26 modifier). A technical unit can be identified by a modifier code of TC. A global unit can be identified by the absence of a TC or 26 modifier code.

SPECT MPI, stress echocardiography, and stress MR studies can be billed separately for the technical and professional components or billed globally, which includes both the professional and technical components.

Professional component claims will outnumber technical component claims due to over-reads.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Studies are excluded for any patients with diagnosis codes in at least three of the following categories: diabetes mellitus, renal insufficiency, stroke or transient ischemic attack, prior heart failure, or ischemic heart disease.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Studies are excluded for any patients with diagnosis codes in at least three of the following categories:

Diabetes (look back of one year)
Diabetes mellitus
ICD-9 codes 249, 250, and 648.0X
ICD-10 codes E08.00-E13.9
Diabetes mellitus in pregnancy, childbirth, and the puerperium
ICD-10 codes O24.011-O24.33, O24.811-O24.93

Renal Insufficiency (look back of one year)
Renal insufficiency

ICD-9 codes 403, 404, 580, 582, 583, 584, 585, 586, and 593.9
Hypertensive chronic kidney disease
ICD-10 codes I12.0-I12.9
Hypertensive heart and chronic kidney disease
ICD-10 codes I13.0-I13.2
Glomerular diseases
ICD-10 codes N00.0-N01.9, N03.0-N03.9, N05.0-N08
Acute kidney failure and chronic kidney disease
ICD-10 codes N17.0-N19
Other disorders of kidney and ureter
ICD-10 codes N28.9-N29

Stroke or transient ischemic attack (look back of three years)
ICD-9 codes 430, 431, 432, 433, 434, 435, 436, 437, 438, 674.0X, and 997.02
Transient cerebral ischemic attacks and related syndromes
ICD-10 codes G45.0-G45.2, G45.8-G45.9
Vascular syndromes of brain in cerebrovascular diseases
ICD-10 codes G46.0-G46.2
Cerebrovascular diseases
ICD-10 codes I60.00-I63.9, I65.21-I65.29, I66.01-I66.9, I67.1, I67.841-I67.89, I69.00-I69.998
Diseases of the circulatory system complicating pregnancy, childbirth and the puerperium
ICD-10 codes O99.411-O99.43

Prior heart failure (look back of three years)
Prior heart failure
ICD-9 codes 425, 428, and 429
Other forms of heart disease
ICD-10 codes I42.0-I43
Heart failure
ICD-10 codes I50.1-I50.9
Intraoperative and post-procedural complications and disorders of circulatory system, not elsewhere classified
ICD-10 codes I97.0-I97.191
Complications and ill-defined descriptions of heart disease
ICD-10 codes I51.0-I51.9

Ischemic heart disease (look back of three years)
Ischemic heart disease
ICD-9 codes 410, 411, 412, 413, and 414
ICD-10 codes I20.0-I22.9, I24.8-I25.119, I25.700-I25.799

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not applicable; this measure does not stratify its results.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable; this measure does not risk adjust.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate

worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
No risk model specifications are provided, as risk adjustment or stratification are not necessary for this measure.

S.16. **Type of score:**
Other (specify):
If other: Percentage

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
This measure calculates the percentage of SPECT MPI, stress echocardiography, or stress MR studies that are performed within the 30 days preceding a non-cardiac, low-risk surgery, out of all SPECT MPI, stress echocardiography, and stress MR studies performed. The measure is calculated based on one year of hospital outpatient claims data, as follows:

1. Select hospital outpatient claims with a CPT code for any SPECT MPI, stress echocardiography, or stress MR on a revenue line item
2. Exclude professional component only claims with modifier =´26´
3. Exclude cases with three or more exclusion diagnoses occurring during the look back period for each diagnosis
4. Set denominator counter = 1
5. Set numerator counter = 1 if a non-cardiac, low-risk surgery occurs within the 30 days following the SPECT MPI, stress echocardiography, or stress MR from step 1, above
6. Aggregate denominator and numerator counts by Medicare provider number
7. Measure = numerator counts / denominator counts [The value should be recorded as a percentage]

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
This measure relies exclusively on 100 percent Medicare fee-for-service (FFS) standard analytical file (SAF) data; no sampling of beneficiaries was performed.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
This measure does not use survey data.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
The measure does not make any adjustments for missing data. The measure relies on Medicare claims data, which are used for payment purposes for services rendered by a provider. The data undergo prepayment claims analysis and post payment audits, as part of the CMS administrative process. The analytic files used by the measure developer are post-adjudicated claims.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database,*

clinical registry, collection instrument, etc.)
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure was initially constructed using the 100-percent FFS outpatient standard analytical files (SAFs) from 2009. These outpatient SAFs contain the claims data on imaging utilization and low-risk surgical procedures performed in hospital outpatient departments (including emergency department services), which are necessary to attribute the measure to specific facilities. Public reporting of the measure currently uses the 100 percent Medicare FFS outpatients SAFs from 2013 and 2014.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospital/Acute Care Facility, Imaging Facility
If other:

S.28. **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; this is not a composite measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_0669_Measure_Testing_Form_2015-06-30.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0669
Measure Title: Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac, Low-Risk Surgery
Date of Submission: 6/30/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/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*).

We tested the measure using 2010 - 2013 Medicare fee-for-service (FFS) data from the 100% samples from the Outpatient Standard Analytic File (SAF-O), Inpatient Standard Analytic File (SAF-I), and Carrier File.

a. Datasets used to define the initial patient population (denominator):
   - *SAF-O:* The initial patient population was defined based on the 2013 100% SAF-O file. The initial patient population includes all claims for a cardiac imaging study from January 1, 2013 – December 1, 2013, provided in a hospital outpatient setting. This dataset also includes unique patient and facility identifiers.
   - *Enrollment database and denominator files:* This dataset contains Medicare FFS enrollment, demographic, and death information for patients identified in the above file.
   - *Provider of services (POS) file:* The POS file contains data on facility characteristics including urbanicity, bed count, and teaching status.

b. Datasets used to capture the numerator:
   - *SAF-O and Carrier:* For patients included in the initial patient population, numerator cases are identified by searching the 2013 100% SAF-O and Carrier files for one or more claims for low-risk, non-cardiac surgery in the 30 days following the cardiac imaging study.

c. Datasets used to identify measure exclusions:

- *SAF-O, SAF-I, and Carrier:* For patients included in the initial patient population, denominator exclusions are identified by searching the 2010 – 2013 100% SAF-O, SAF-I, and Carrier files for risk factor diagnoses in the three years preceding the cardiac imaging study.

1.3. What are the dates of the data used in testing? 2010-2013

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.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: state, national	☒ other: state, national

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*)
The number of measured entities (hospital outpatient departments) varies by testing type; see Section **1.7** for details.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
The number of patients varies by testing type; see Section **1.7** for details.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.
The data sources, dates, number of measured entities, number of cardiac imaging studies, number of surgeries, level of analysis, and demographic profile for the patients used in each type of testing are as follows:

Reliability Testing
Data Source: Denominator: SAF-O; Numerator: SAF-O and Carrier; Exclusions: SAF-O, SAF-I, and Carrier
Dates: Denominator: January 1, 2013 – December 1, 2013; Numerator: January 1, 2013 – December 31, 2013; Exclusions: January 1, 2010 – December 1, 2013
Number of Facilities: 2,759
Number of Cardiac Imaging Studies: 1,042,496
Number of Surgeries: 53,579
Level of Analysis: Facility
Patient Characteristics: Gender (% Male): 46.4; Median Age (Years): 70.3 (St. Dev.: 10.2); Race/Ethnicity (% Minority): 14.9

Validity Testing
Data Source: Structured qualitative survey questions completed by Technical Expert Panel members

Date Collected: June 2015
Number of Responses: 7
Respondent Characteristics: Respondents were asked to select at least one of the following categories: insurer/purchaser (2); clinician (4); management/administration (2); patient/patient advocate/caregiver (3).

Exclusions Analysis
Data Source: Denominator: SAF-O; Numerator: SAF-O and Carrier; Exclusions: SAF-O, SAF-I, and Carrier
Dates: Denominator: January 1, 2013 – December 1, 2013; Numerator: January 1, 2013 – December 31, 2013; Exclusions: January 1, 2010 – December 1, 2013
Number of Facilities: 2,770
Number of Cardiac Imaging Studies: 1,043,534
Number of Surgeries: 53,824
Level of Analysis: Observation, Facility
Patient Characteristics: Gender (% Male): 46.4; Median Age (Years): 70.2 (St. Dev.: 10.2); Race/Ethnicity (% Minority): 14.9

Risk Adjustment/Stratification
N/A

Identification of Statistically Significant & Meaningful Differences in Performance
Data Source: Denominator: SAF-O; Numerator: SAF-O and Carrier; Exclusions: SAF-O, SAF-I, and Carrier
Dates: Denominator: January 1, 2013 – December 1, 2013; Numerator: January 1, 2013 – December 31, 2013; Exclusions: January 1, 2010 – December 1, 2013
Number of Facilities: 2,759
Number of Cardiac Imaging Studies: 1,042,496
Number of Surgeries: 53,579
Level of Analysis: Facility
Patient Characteristics: Gender (% Male): 46.4; Median Age (Years): 70.3 (St. Dev.: 10.2); Race/Ethnicity (% Minority): 14.9

Comparability of Performance Scores when more than one Set of Specifications
N/A

Missing Data Analysis and Minimizing Bias
N/A

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)
Reliability was calculated using two tests. We first assessed the ability of the measure to identify statistical outliers. As a secondary assessment, we calculate reliability in accordance with the methods discussed in *The Reliability of Provider Profiling: A Tutorial* (2009). The reliability testing calculates the ability of the measure to distinguish between the performance of different facilities. Specifically, the testing calculated the signal-to-noise ratio for each facility meeting the minimum case count in 2013. The reliability score is estimated using a beta-binomial model, which is appropriate for the reliability testing of pass/fail measures. The reliability score for each facility is a function of the facility's sample size and score on the measure, and the variance across facilities.

Reference:
Adams JL. The reliability of provider profiling: a tutorial. Santa Monica, CA: RAND Corporation. 2009. Retrieved from http://www.rand.org/pubs/technical_reports/TR653.

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*)
Of the 2,759 facilities in 2013 meeting the minimum case count, 137 (5.0 percent) facilities had a performance value that was statistically significantly different from the weighted mean (or benchmark value). Statistically meaningful difference was defined as when the facility score fell outside of the confidence interval (± 1.96 standard deviations) for the measure mean (benchmark value).

Figure 1 (below) is a histogram of the distribution of the reliability scores for the facilities meeting the minimum case count in 2013. Reliability scores ranged from 4.9 percent to 100.0 percent, with a median reliability score of 43.0 percent.

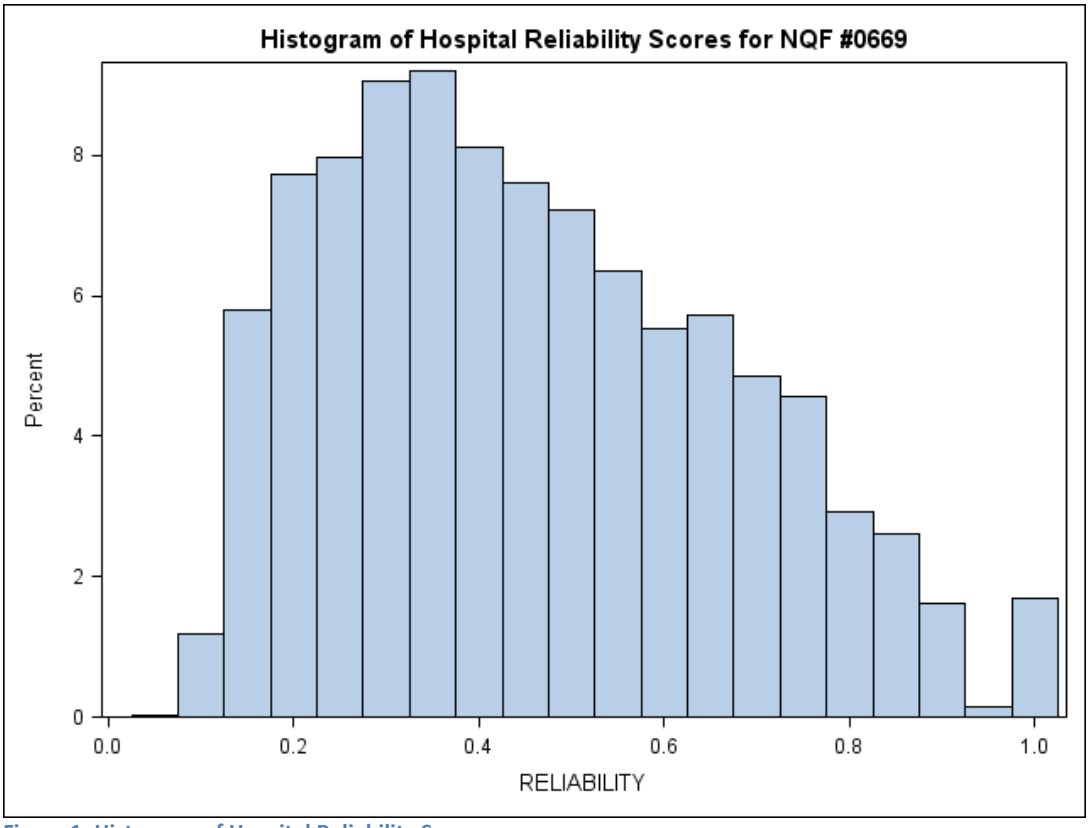


Figure 1: Histogram of Hospital Reliability Scores

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?*)
As the intent of the measure is to identify differences from the mean (or threshold value), the ability of the measure to identify statistical outliers is indicative of strong reliability. As noted above, 137 facilities were classified as outliers in 2013, indicating that these facilities had a rate of overuse that was statistically significant and meaningfully different from the measure mean.

Calculated using a beta-binomial model, a median reliability score of 43.0 percent is indicative of moderate measure reliability and further supports the findings of the test to identify statistical outliers. A beta-binomial model is intended to assess the ability of the measure to identify true differences in performance between individual facilities.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*
Face validity of the measure score was systematically assessed through survey of the Technical Expert Panel (TEP). Seven TEP members participated in the data collection. Respondent perspectives include insurers/purchasers, clinicians, management or administration, patients/patient advocates, and caregivers. Prior to responding to questions related to measure-score and data-element face validity, TEP members were provided detailed measure specifications.

The following questions and statements related to measure-score face validity were posed to the TEP:
1. To gather data, the measure uses claims to look 30 days forward from the date of the cardiac imaging test to determine if the patient had a low-risk surgery. Thirty days is an appropriate time period to capture a surgery related to the preoperative imaging test.
2. This measure helps assess the inappropriate use of pre-operative cardiac imaging.

Data-element face validity was also assessed, using the following questions and statements:
1. The following tests can be accurately captured using claims data: stress echocardiography, SPECT MPI, and stress MRI.
2. Do you foresee any challenges in capturing any of these exclusions in claims data?
3. For this measure, indications for measure exclusion include any patients with diagnosis codes from the following categories: diabetes mellitus, renal insufficiency, stroke or transient ischemic attack, prior heart failure, or ischemic heart disease. Please indicate how well the following patient exclusions identify high-risk patients.

Responses to questions *1* and *2* in the measure-score face-validity section and question *1* in the data-element face-validity section were collected using a five-point scale: strongly agree, agree, undecided, disagree, strongly disagree, and do not know. For data-element face validity, responses to question *2* were collected using yes/no response options; responses to question *3* were collected using keep/remove response options.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Results of the face-validity assessment indicate that a diverse group of stakeholders, a majority of whom were not involved in the measure's development, support the validity of the measure. Results for each of the questions provided above follow.

Measure-Score Face Validity

1. *To gather data, the measure uses claims to look 30 days forward from the date of the cardiac imaging test to determine if the patient had a low-risk surgery. Thirty days is an appropriate time period to capture a surgery related to the preoperative imaging test.*

Response Option	Response Percentage	Response Count
Strongly Agree	42.9%	3
Agree	28.6%	2
Undecided	0.0%	0
Disagree	0.0%	0
Strongly Disagree	0.0%	0
Do Not Know or Not Applicable	28.6%	2

2. *This measure helps assess the inappropriate use of pre-operative cardiac imaging.*

Response Option	Response Percentage	Response Count
Strongly Agree	42.9%	3
Agree	28.6%	2
Undecided	14.3%	1
Disagree	0.0%	0
Strongly Disagree	0.0%	0
Do Not Know or Not Applicable	14.3%	1

Data-Element Face Validity

1a. The following test can be accurately captured using claims data: stress echocardiography

Response Option	Response Percentage	Response Count
Strongly Agree	42.9%	3
Agree	42.9%	3
Undecided	0.0%	0
Disagree	0.0%	0
Strongly Disagree	0.0%	0
Do Not Know or Not Applicable	14.3%	1

1b. The following test can be accurately captured using claims data: SPECT MPI

Response Option	Response Percentage	Response Count
Strongly Agree	42.9%	3
Agree	42.9%	3
Undecided	0.0%	0

Disagree	0.0%	0
Strongly Disagree	0.0%	0
Do Not Know or Not Applicable	14.3%	1

1c. The following test can be accurately captured using claims data: stress MRI

Response Option	Response Percentage	Response Count
Strongly Agree	42.9%	3
Agree	42.9%	3
Undecided	0.0%	0
Disagree	0.0%	0
Strongly Disagree	0.0%	0
Do Not Know or Not Applicable	14.3%	1

2. *Do you foresee any challenges in capturing any of these exclusions in claims data?*

Response Option	Response Percentage	Response Count
No	71.4%	5
Not Sure/ Do Not Know	14.3%	1
Yes (please explain)	14.3%	1: ICD-9 is not robust enough to capture all of the clinical nuances that would be required to differentiate which of these patients should be excluded.

3. *For this measure, indications for measure exclusion include any patients with diagnosis codes from the following categories: diabetes mellitus, renal insufficiency, stroke or transient ischemic attack, prior heart failure, or ischemic heart disease. Please indicate how well the following patient exclusions identify high-risk patients.*

Exclusion	Keep this Exclusion	Remove this Exclusion	Do Not Know/ Not Applicable
Diabetes mellitus	2	2	3
Renal insufficiency	3	1	3
Stroke or transient ischemic attack	3	0	4
Prior heart failure	3	0	4
Ischemic heart disease	3	0	4

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The face validity of this measure was evaluated by our Technical Expert Panel. Results for four of the questions listed above indicate that the measure is appropriately specified. Seventy-one percent of respondents believe the 30-day window used to look forward for a low-risk, non-cardiac surgery from the date of the imaging procedure accurately captures imaging related to pre-surgical testing. Eighty-six percent of those who responded believe

stress echocardiography, SPECT MPI, and stress MRI are accurately captured using claims data. Seventy-one percent of respondents believe the exclusions are accurately captured using claims data.

For one question listed above, no consensus was reached related to which clinical conditions should be excluded from the measure; this lack of consensus was driven by *do not know* responses. For both this question (and for others for which a response of *Do Not Know* was selected), respondents did not have feel they had the clinical knowledge to provide a definitive response. Though TEP members did not reach consensus for this question, related to the clinical conditions to be excluded from the measure, evidence from the literature and data included in Section **2b3** support continued exclusion of these five clinical categories. The European Society of Cardiology recommends stress testing only in patients with three or more clinical risk factors for cardiac complications after surgery, in accordance with Lee Index preoperative risk stratification model.

Historically, measures that rely on claims data for calculation of performance are assumed to have strong face validity. This assumption of face validity is due in part to the rigor with which data are cleaned and audited prior to payment and subsequent use in measure calculation. For other public reporting programs for which payment is adjusted based on provider performance, few concerns about use of claims data for face validity have been raised.

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

We tested measure exclusions to determine the prevalence of each exclusion, by facility, and at an aggregate level. We also tested the aggregate risk factor exclusion to determine the effect on facility performance score, both by reporting summary statistics and by calculating a spearman rank correlation coefficient. The analysis tested the following categories of measure exclusions using 2013 performance data:

- Diabetes
- Renal insufficiency
- Stroke or transient ischemic attack
- Prior heart failure
- Ischemic heart disease
- Three or more of the above risk factors

Currently, the measure excludes patients with three or more of the listed risk factors in the prior three years, in accordance with the Lee Index (European Society of Cardiology, 2009).

Reference:

European Society of Cardiologists. Guidelines for pre-operative cardiac risk assessment and perioperative cardiac management in non-cardiac surgery. Eur Heart J. 2009; 30: 2769–812.

2b3.2. 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*)

We examined overall frequencies and proportions of the studies excluded for each exclusion criterion in all cardiac imaging studies for a sample of 2,770 facilities meeting the minimum case count in 2013 if no exclusions were imposed. The initial patient population included 647,957 cardiac imaging studies. The final initial patient population included 633,195 cardiac imaging studies. The total number of exclusion occurrences exceeded the number of studies excluded because a single patient might meet multiple exclusion criteria.

Exclusion	Overall Occurrence (N)	Overall Occurrence (%)	Distribution Across Facilities (%)		
			25th	50th	75th
Diabetes	461,534	44.2	38.6	43.8	50.0
Renal Insufficiency	292,409	28.0	22.9	27.2	31.9
Stroke or Transient Ischemic Attack	338,431	32.4	25.9	30.9	36.6
Prior Heart Failure	449,600	43.1	35.1	41.7	49.2
Ischemic Heart Disease	643,962	61.7	50.9	58.7	67.2
Any three of the risk factors	395,577	37.9	29.9	36.4	43.6

Additionally, we calculated descriptive statistics for the measure scores of each facility, with and without the exclusion for patients with three or more risk factors.

Descriptive Statistic	With Exclusions (%)	Without Exclusions (%)
Minimum	0.0	0.0
Maximum	22.1	23.0
Mean	4.9	5.1
Standard Deviation	2.3	2.2
25th Percentile	3.4	3.7
50th Percentile (Median)	4.7	4.9
75th Percentile	6.2	6.3

Finally, we calculated a spearman rank correlation coefficient for facility score with and without the exclusion for patients with three or more risk factors:

$r_S = 0.8304$
$p = 0.0000$

2b3.3. 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.* <u>*Note*</u>*:* ***If patient preference is an exclusion****, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The frequency of the exclusion for patients with three or more risk factors varied substantially across facilities (IQR: 13.7 percent). While median performance does not change significantly, exclusion of these patients is appropriate, as inclusion would noticeably alter the initial patient population for select facilities. As noted in the 2009 European Society of Cardiology guideline, patients risk should be determined by evaluating clinical risk factors, test results and intensity of the planned surgical procedure to arrive at an individualized cardiac risk assessment. Using the six-variable Lee Index preoperative risk stratification model, the European Society of Cardiology recommends stress testing only in patients with three or more clinical risk factors for cardiac complications after surgery. The six clinical risk factors used for preoperative cardiac risk stratification in the Lee Index model are:

- IHD (angina pectoris and/or MI)
- Surgical risk: high-risk surgery
- Heart failure
- Stroke/transient ischemic attack
- Diabetes mellitus requiring insulin therapy
- Renal dysfunction/hemodialysis

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use 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.

This measure is a process measure for which we provide no risk adjustment or risk stratification. We determined risk adjustment and risk stratification were not appropriate based on the measure evidence base and the measure construct. During the measure development and maintenance process, we performed an annual review of the literature, which included a scan for potential patient subpopulations for which there are differences in the clinical decision to perform cardiac imaging prior to low-risk, non-cardiac surgery; this review identified no clear evidence of an empirical relationship between sociodemographic status (SDS) and facility-level measure performance.

In addition to the evidence gathered from the literature, stakeholder feedback obtained during the four years of implementation and public reporting has not identified concerns related to SDS factors and need for risk adjustment. This supports the conceptual model upon which the measure is based. As a process-of-care measure, the decision to image a patient prior to surgery should not be influenced by SDS factors; rather, adjustment would risk masking such important inequities in care delivery. Variation across patient populations is reflective of differences in the quality of care provided to the disparate patient population included in the measure's denominator.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

No risk adjustment or stratification was performed.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

No risk adjustment or stratification was performed.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

No risk adjustment or stratification was performed.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)
Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
No risk adjustment or stratification was performed.

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
No risk adjustment or stratification was performed.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
No risk adjustment or stratification was performed.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
No risk adjustment or stratification was performed.

2b4.9. Results of Risk Stratification Analysis:
No risk adjustment or stratification was performed.

2b4.10. 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*)
No risk adjustment or stratification was performed.

2b4.11. 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*)
No risk adjustment or stratification was performed.

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*
The measure calculation contractor takes a number of steps to ensure precision and accuracy of publicly reported values. One step in this process is through application of a minimum case count to exclude facilities that do not perform a high volume of services contained within the measure's specifications. In the situation where a facility provides only a handful of the relevant services that are eligible for this measure, the results of the measure may be significantly impacted and skewed by one or two cases. Minimum case count requirements were developed for each facility in order to assure a 90 percent confidence level for the observed facility rate. There are two different processes for determining required case counts depending on whether the facility rate is less than 0.05 or greater than 0.95 (i.e., towards the end of the range of possible rate values), or somewhere between 0.05 and 0.95 (inclusive). Each process has three steps: (1) determine reasonable levels of precision; (2) determine the level of confidence to be required for the measures; and, (3) calculate the case counts needed to meet the precision requirements. For facility rates less than 0.05 or greater than 0.95, the case count needed to attain the required precision was calculated to be 45 cases. For facility rates between 0.05 and 0.95, the case count needed to attain the required precision ranges from 31 to 67 cases. For more details on the minimum case count requirements determinations, please see the supplemental materials.

Following the application of the minimum case count, we also tested the statistical significance of the difference between facility performance scores and the mean performance value. For the 2013 data, this included 2,383 facilities. For each facility, the facility performance score and standard deviation was calculated. This analysis identified more than 130 facilities as statistical outliers. Additional details of this analysis are provided in Section **2b5.2**.

Methodology explaining the minimum case count calculations for this measure can be found at https://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228889854907&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3D2012_OIE_MCC.pdf&blobcol=urldata&blobtable=MungoBlobs.

2b5.2. 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*)
Of the 2,759 facilities in 2013 meeting the minimum case count, 137 (5.0 percent) facilities had a performance value that was statistically significantly different from the weighted mean (or benchmark value). Statistically meaningful difference was defined as when the facility score fell outside of the confidence interval (± 1.96 standard deviations) for the measure mean (benchmark value). Thus, this calculation identifies statistical outliers.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
Analysis of the 2013 performance data, and the subsequent rate of identification of statistically different performance for 7.4 percent of measured entities, demonstrates the ability of the measure to identify outlying performance. While many facilities have converged around a performance score of between 4 and 5 percent, more than 7 percent of facilities continue to have outlying performance. By reporting a measure mean (benchmark value), this provides an opportunity for outlying facilities to identify their high rate of overuse and work to implement quality improvement mechanisms to reduce the rate of overuse of cardiac imaging studies.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)
This measure only uses one set of specifications.

2b6.2. 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*)
This measure only uses one set of specifications.

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)
This measure only uses one set of specifications.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)
This measure is calculated from claims data submitted by facilities for purposes of payment. The administrative claims data used to calculate the measure are maintained by CMS's Office of Information Services; these data undergo additional quality assurance checks during measure development and maintenance. Thus, the analytic files used for measure testing and measure calculation include post-adjudicated claims, and do not include missing data.

2b7.2. 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*)
As described in Section **2b7.1**, the analytic files used for measure testing and measure calculation include post-adjudicated claims, and do not include missing data.

2b7.3. 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*)
As described in Section **2b7.1**, the analytic files used for measure testing and measure calculation include post-adjudicated claims, and do not include missing data. As such, missing data does not bias the 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.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources
The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to

electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields? *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.

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.
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. Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
This measure is claims based and uses CMS hospital outpatient claims as its data source.

Special attention needs to be taken when counting procedures on the Medicare claims files. The biggest issue is how to deal with modifier codes. Modifiers are two digit indicators (alpha or numeric) that represent a service or procedure that has been altered by some specific circumstance, which typically will impact the payment amount.

Procedure modifier code “26” represents the professional component of a procedure and includes the clinician work (i.e., the reading of the image by a physician), associated overhead and professional liability insurance costs. This modifier corresponds to the human involvement in a given service or procedure.

The procedure modifier code “TC” represents the technical component of a service or procedure and includes the cost of equipment and supplies to perform that service or procedure. This modifier corresponds to the equipment/facility part of a given service or procedure.

In most cases, unmodified codes represent a global procedure which includes both the professional and technical components. There are also other modifier codes. All other modifier codes have been counted as a technical code for our purposes. When calculating the measures, we are only concerned with procedures associated with technical and global modifiers, as these modifiers refer to services provided by the facility. This reduces the possibility of double-counting procedures, since a single procedure may result in both a technical and professional record on the claims files. There were very few instances when this occurred as it related to procedures applicable to the measure.

When developing counts of procedures, the objective is to avoid double-counting procedures that may have been billed through multiple revenue centers within a facility. Billing through multiple centers leads to multiple records in the Medicare claims files (i.e., the SAFs). For instance, there may be multiple bills for a single SPECT MPI. On one bill, the charges relate to the application of a radiopharmaceutical, which could have a technical modifier code and come from the pharmacy revenue center. On the other bill, the charges relate to the imaging study and may fall under a technical bill from the imaging center revenue center. In this case, we only count the SPECT MPI once, since only one SPECT MPI was performed. However, if we were summing up the Medicare paid amounts for this procedure, we would include the Medicare paid amounts from both bills, as they each represent payments for

services directly related to the particular SPECT MPI procedure.

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)*.
No fees, licensure, or other requirements are necessary to use this measure; however, CPT codes, descriptions, and other data are copyright 2013 American Medical Association. All rights reserved. CPT is a registered trademark of the American Medical Association. Applicable FARS\DFARS Restrictions Apply to Government Use. Fee schedules, relative value units, conversion factors, and/or related components are not assigned by the AMA, are not part of CPT, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein.

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.

Planned	Current Use (for current use provide URL)
	Public Reporting Hospital Outpatient Quality Reporting http://www.medicare.gov/hospitalcompare/search.html Hospital Outpatient Quality Reporting https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228695266120 Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Hospital Outpatient Quality Reporting http://www.medicare.gov/hospitalcompare/search.html Hospital Outpatient Quality Reporting https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228695266120

4a.1. **For each CURRENT use, checked above, provide:**

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

Public Reporting:
Name of program and sponsor: The CMS Hospital Outpatient Quality Reporting (HOQR) Program
Purpose: The HOQR Program is a pay for quality data reporting program implemented by CMS for outpatient hospital services. In addition to providing hospitals with a financial incentive to report their quality of care measure data, the HOQR Program provides CMS with data to help Medicare beneficiaries make more informed decisions about their health care. Hospital quality of care information gathered through the HOQR Program is publicly available on the Hospital Compare website.
Accountable entities and patients: The publicly reported values (on Hospital Compare) are calculated for all facilities in the United States that meet minimum case count requirements. For the period of 2009 to 2014, 1,953 facilities met the minimum case count each year. Additional facilities met the minimum case count requirements in some, but not all, years. The claims included in the publicly reported calculations are for Medicare FFS patients whose claims are subject to the Outpatient Prospective Payment System

(OPPS).

Quality Improvement with Benchmarking (external benchmarking to multiple organizations):
Name of program and sponsor: The CMS HOQR Program
Purpose: The HOQR Program is a pay for quality data reporting program implemented by CMS for outpatient hospital services. In addition to providing hospitals with a financial incentive to report their quality of care measure data, the data is publicly reported on the Hospital Compare Website. The data reported on Hospital Compare not only shows the hospital's score on the measure, but also provides state and national averages for the measure. This enables consumers to compare the hospital's performance to other facilities and determine if the facility is an outlier.
Accountable entities and patients: The publicly reported values (on Hospital Compare) are calculated for all facilities in the United States that meet minimum case count requirements. For the period of 2009 to 2014, 1,953 facilities met the minimum case count each year. Additional facilities met the minimum case count requirements in some, but not all, years. The claims included in the publicly reported calculations are for Medicare FFS patients whose claims are subject to the OPPS.

4a.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?*)
This measure is publicly reported.

4a.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.*)
This measure is publicly reported.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Summary statistics for performance scores from 2009 - 2014 are provided in Section 1b.2.

The median rate of overuse increased from 2009 to 2011 (4.92 percent to 5.50 percent); however, the rate of overuse has declined since that point. From July 2013 – June 2014 the median rate of overuse fell to 4.88 percent. Over the period of January 2009 – June 2014, 1,953 facilities met the minimum case count to be eligible for public reporting in all years. Additional facilities met the minimum case count requirements in some, but not all, years. During the July 2012-June 2013 reporting period, there were more than 45,000 inappropriate cardiac imaging studies performed for Medicare FFS beneficiaries. This number fell to approximately 28,000 potentially inappropriate imaging studies in the most recent year of publicly reported data, July 2013-June 2014.

4b.2. **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.**
Not applicable as there is demonstrated improvement in measure performance.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We did not identify any unintended consequences during measure testing. Similarly, no evidence of unintended consequences to

individuals or populations have been reported since implementation. We will continue to monitor the potential for unintended consequences through an annual review of the literature as well as an ongoing review of stakeholder comments and inquiries.

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)**
0670 : Cardiac stress imaging not meeting appropriate use criteria: Preoperative evaluation in low risk surgery patients

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
Institute for Clinical Systems Improvement (ICSI) - Perioperative protocol: percentage of patients undergoing elective non-high-risk surgery having laboratory tests/imaging unrelated to positive findings on preoperative basic health assessment

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Although NQF #0669 is similar to NQF #0670, there are several differences that would make measure harmonization infeasible and reduce the effectiveness of both currently endorsed measures. First, the measures serve different target populations and purposes: the CMS measure is used for public reporting and the measure calculations only include CMS FFS claims; on the other hand, the ACC measure is not restricted to the Medicare population and the measure calculations are sold to hospitals as part of a quality improvement package, rather than used for public reporting. Second, the measures include different stress testing procedures: the ACC measure (NQF #0670) includes SPECT MPI, stress echocardiography, CCTA, and CMR procedures codes in the denominator, whereas the CMS measure (NQF #0669) includes SPECT MPI, stress echocardiography, and stress MR procedure codes. Finally, the ACC measure relies on a different data source than does the CMS measure: unlike the CMS measure, the ACC measure does not account for instances where the imaging and low risk surgery occur at different facilities. While NQF #0669 is related to the ICSI measure, significant structural differences makes measure harmonization inappropriate for these measures. The denominator of the ICSI measure is defined by low-risk surgery cases, whereas the denominator of the CMS measure is defined by cardiac imaging studies. The ICSI measure also relies on test results for measure calculation, a data element not available in CMS administrative claims data. Finally, the ICSI measure includes patients aged 2 years and older while the CMS measure is targeted to the Medicare population.

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.)
We did not identify any competing measures that address both the same measure focus and target population as NQF #0669.

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):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Vinitha, Meyyur, Vinitha.Meyyur@cms.hhs.gov, 410-786-7224-
Co.3 **Measure Developer if different from Measure Steward:** The Lewin Group
Co.4 **Point of Contact:** Colleen, McKiernan, Colleen.McKiernan@lewin.com, 703-269-5595-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The contractor has convened a Technical Expert Panel (TEP), which will evaluate and provide feedback on measure-development and maintenance efforts for the imaging efficiency measures. Specifically, the TEP will provide direction and feedback through all phases of project activities, including expansion of imaging efficiency measures to additional CMS quality reporting programs, updates to the current specifications of the seven imaging efficiency measures, review of quantitative testing results, feedback on qualitative testing questions (i.e., results of TEP member questionnaires), and support for endorsement of the measures by the National Quality Forum (NQF).

The following is a list of the contractor's TEP members:

Meenu Arora, MBA
Quality Improvement Leader , Sequoia Hospital

Brian Baker
Chief Executive Officer, Carealytics

Peter Benner
Vice Chair, MNSure

Martha Deed, Ph.D
Safe Patient Project´s Patient Advocacy Network

Lawrence Feinberg, MD
Attending Physician, University of Colorado Hospital

Elliott Fishman, MD
Professor of Radiology and Oncology, Johns Hopkins School of Medicine

Marian Hollingsworth
Patient Advocate

Michael Hutchinson, MD Ph.D

Clinical Associate Professor of Neurology, Icahn School of Medicine at Mount Sinai

Gregory M. Kusiak, MBA FRBMA
President, California Medical Business Services, Inc.

Barbara Landreth, RN MBA
Clinical Information Analyst , St. Louis Area Business Health Coalition

Barbara McNeil, MD Ph.D
Head Professor of Radiology, Harvard University

Michael J. Pentecost, MD
Chief Medical Officer, NIA Magellan

David Seidenwurm, MD
Medical Staff Consultant, Sutter Medical Group

Adam Sharp, MD MS
Research Scientist , Kaiser Permanente Southern California

Paul R. Sierzenski, MD RDMS FACEP FAAEM
Medical Director, Christian Health Care System

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2011

Ad.3 **Month and Year of most recent revision:** 03, 2015

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

Ad.5 **When is the next scheduled review/update for this measure?** 03, 2016

Ad.6 **Copyright statement:** This measure does not have a copyright.

Ad.7 **Disclaimers:** CPT codes, descriptions, and other data only are copyright 2013 American Medical Association. All rights reserved. CPT is a registered trademark of the American Medical Association. Applicable FARS\DFARS Restrictions Apply to Government Use. Fee schedules, relative value units, conversion factors and/or related components are not assigned by the AMA, are not part of CPT, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein.

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET - COMPOSITE

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 #: 0694
Measure Title: Hospital Risk-Standardized Complication Rate following Implantation of Implantable Cardioverter-Defibrillator (ICD)
Measure Steward: American College of Cardiology
Brief Description of Measure: This measure provides hospital specific risk-standardized rates of procedural complications following the implantation of an ICD in patients at least 65 years of age. The measure uses clinical data available in the National Cardiovascular Data Registry (NCDR) ICD Registry for risk adjustment linked with administrative claims data using indirect patient identifiers to identify procedural complications.
Developer Rationale: The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized procedural complication rates following hospitalization for an ICD implantation. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as operator and hospital procedural expertise, communication among providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This measure was developed to identify institutions whose performance is better or worse than would be expected based on their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

Numerator Statement: The outcome for this measure is one or more complications within 30 or 90 days (depending on the complication) following initial ICD implantation. The measure treats complications as a dichotomous (yes/no) variable; we are interested in whether or not a complication has occurred and not how many complications occurred in each hospital.
Denominator Statement: The target population for this measure includes inpatient and outpatient hospital stays with ICD implants for patients at least 65 years of age who have matching information in the National Cardiovascular Disease Registry (NCDR) ICD Registry. The time window can be specified from one to three years. This measure was developed with Medicare claims and CathPCI Registry data from one calendar year (2007).
Denominator Exclusions: (1) Previous ICD placement. Hospital stays in which the patient had an ICD implanted prior to the index hospital stay are excluded.
Rationale: Ideally, the measure would include patients with a prior ICD, as this is a population known to be at high risk of adverse outcomes. However, for these patients it is difficult to distinguish in the administrative data whether adverse events such as infection were present on admission or complications of the second ICD placement. In order to avoid misclassification, we exclude these patients from the measure.
(2) Previous pacemaker placement, Hospital stays in which the patient had a previous pacemaker placement prior to the index hospital stay are excluded.
Rationale: Some complications (infection or mechanical complication) may be related to a pacemaker that was removed prior to placement of an ICD. Ideally, the measure would include patients with a prior pacemaker, as this is a population known to be at higher risk of adverse outcomes. However, for these patients it is difficult to distinguish in the administrative data whether adverse events such as infection were present on admission or complications of the ICD placement. In order to avoid misclassification, we exclude these patients from the measure.
(3) Not Medicare FFS patient on admission. Patient admissions in which the patient is not enrolled in Medicare FFS at the time of the ICD procedure.
Rationale: Outcome data are being derived only for Medicare fee-for-service patients.
(4) Lack 90-day follow-up in Medicare FFS post-discharge. Patients who cannot be tracked for 90 days following discharge are excluded.
Rationale: There will not be adequate follow-up data to assess complications
(5) Not the first claim in the same claim bundle. There are cases when several claims in the same hospital representing a single

episode of care exist in the data together. These claims are bundled together and any claim other than the first is excluded. Rationale: Inclusion of additional claims could lead to double counting of an index ICD procedure.

Measure Type: Composite
Data Source: Administrative claims, Electronic Clinical Data : Registry
Level of Analysis: Facility, Population : National
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

1d.1. **Composite Measure Construction:** any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: 1/17/11 **Most Recent Endorsement Date:** 1/17/11

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from National Voluntary Consensus Standards for Patient Outcomes 2009

This measure provides hospital-specific risk-standardized rates of procedural complications following the implantation of an ICD in patients at least 65 years of age. The measure uses clinical data available in the National Cardio- vascular Data Registry (NCDR) ICD Registry for risk-adjustment that has been linked with CMS administrative claims data used to identify procedural complications. This measure can be applied to all Medicare patients at least 65 years of age.

This measure was designed to combine clinical data from the National Cardiovascular Data Registry (NCDR)6 ICD Registry and administrative data. All patients over age 65 years are required to be entered into the registry, and 70 percent of hospitals report all patients to NCDR. The Committee and TAP agreed that the measure is important in addressing a costly procedure that has a high complication rate (18 percent). The TAP also commended the strong performance charac- teristics of the risk model. Committee members were interested in including patients below age 65 years. The measure developers advised the Committee that the measure was developed in the Medicare 65 and older fee-for-service population because this is the only cohort of patients for whom the data are available to reliably identify outcomes (complications and vital status) beyond the index hospitalization. The measure could be applied to a broader population of patients undergoing ICD implantation if the required data elements were available with some additional work to optimize the risk-adjustment methodology.

A Committee member noted that the variation of values in the technical report is very narrow due to hierarchical modeling and therefore will not discriminate among providers. Others suggested that clustering of the complication rate at 18 percent represents opportunity for improvement overall. This measure addresses the National Priority of safety.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing a rationale that supports the relationship of the outcome to at least one healthcare structure, process, intervention, or service.

- This measure calculates hospital-specific risk-standardized rates of procedural complications for patients at least 65 years of age within 30 or 90 days (depending on the complication) following initial ICD implantation.
- The developer provides the following rationale for measuring this outcome:
  - The developer notes that a complication following placement of an Implantable Cardioverter Defibrillator (ICD) is an undesirable outcome, and suggests that complications in the ICD patient population can be reduced through comprehensive, personalized risk assessment, competency of the physician and/or hospital treating the patient and appropriate follow up.

- The developer states that the risk of adverse outcomes following ICD implantation varies markedly by the experience and training of the implanting physician, the device implanted, and the characteristics of the facility in which the procedure is performed.
- The developer also states that this measure was developed to identify institutions whose performance is better or worse than would be expected based on their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality.
- The developer provides a number of references to support their rationale.

Question for the Committee:

- *Does the Committee agree that hospitals have the ability to influence rates of procedural complications following ICD implantation?*

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

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

- The developer states that reported complication rates following ICD implantation vary from 4% to 30%, depending on how complications are defined, the period of assessment and data source.
- The developer reports that in the ACC's NCDR ICD Registry, the incidence of in-hospital complications is approximately 4%, while also noting that complications such as device infection, malfunction, or cardiac tamponade may only become evident following hospital discharge.
- The developer does not provide performance scores for the measure as specified; however, as part of the measure's development, the developer analyzed unadjusted rates of ICD-related complications in 2007 Medicare inpatient claims data, which included 67,532 ICD admissions for 67,080 patients at 1,792 hospitals.
  - In these preliminary analyses, complications were seen in **5.7% of ICD admissions** (3,818 complications); the median complication rate following ICD implantation **ranged from 0% to 17.8% across deciles** of hospitals grouped by their all-cause complication rate.
- The developer examined health disparities associated with this measure by race/ethnicity [percentage of African American patients] and by socioeconomic status (SES) [from AHRQ SES Index data]. For both sets of analyses, the distribution of race and SES were stratified by quintiles.
  - This information was derived from CMS fee-for-service (FFS) data from the time period of 2010Q2 – 2011Q4, comprising 1,279 hospitals and 43,711 patients.
- The developer provides the results of these analyses in the form of box-and-whisker plots that are included in the testing attachment.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

1d. Composite - Quality Construct and Rationale

1c. Composite Quality Construct and Rationale. The quality construct and rationale should be explicitly articulated and logical; a description of how the aggregation and weighting of the components is consistent with the quality construct and rationale also should be explicitly articulated and logical.

- This is an 'any-or-none' composite measure, meaning that the measure treats complications as a dichotomous (yes/no) variable; if any of the specified complications occurs in a given patient, that patient is included in the measure numerator.
- The developer notes that their interest is in whether or not a complication has occurred and not how many

complications occurred in each hospital.

- The developer states that the goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized procedural complication rates following hospitalization for an ICD implantation.

Questions for the Committee:

○ *Are the quality construct and rationale for the composite explicitly stated and logical?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - Acknowledge that measure was developed with Medicare claims and CathPCI registry data, but this measure has limits to its evidence based only on evaluating patients 65 years of age or older. Exclusion criteria are logical except for the issue of excluding all patients who are not Medicare FFS based on the outcome data being derived from Medicare FFS patients. Also, patients who cannot be tracked for 90 days are excluded which could be an issue as patients could have an a complication even if not followed for 90 days. It is an important measure to evaluate quality within hospitals, but is limited by its scope focused on patients 65 years of age or older based on available data.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- This is an any-or-non composite outcome measure: Any complication within 30 or 90 days of a first ICD implantation.
- Evidence presented from 2007 Medicare inpatient claims showed a 5.7% complication rate in ICD admissions with a range of 0% to 17.8% which indicates a huge range of complication rates amongst hospitals and a reason to evaluate this measure.

1b. ***Committee's Comments on Performance Gap:***

- Performance analysis shows that RSCRs vary from about 3% to 14%
- There is a performance gap with the range of complications being as high as 18% in some hospitals and as low as 0% in other with a average of 5.7% overall. When looking at SES and AA race there was no difference in average complication rates across hospitals.

1c. ***Committee's Comments on Composite Performance Measure:***

- The quality construct is logical. An any-or-none composite measure makes differences among hospitals much more clear while each single complication may not draw attention to differences in performance.
- The construct is appropriate based on looking whether a complication occurred in individual patients vs looking at how many overall complications occurred. This allows for evaluation to compare hospital to hospital and the overall mean complication rate.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- This measure calculates complications within 30 or 90 days (depending on the complication) following initial ICD implantation.
- The measure employs a hierarchical logistic regression model to create a hospital-level 30 or 90 day risk-standardized complication rate (RSCR).
- The measure cohort includes inpatient and outpatient hospital stays with ICD implants for patients at least 65 years of age who have matching information in the National Cardiovascular Disease Registry (NCDR) ICD

Registry.

- The denominator is defined using ICD-9 procedure codes from inpatient claims and HCPCS/CPT procedure codes from outpatient claims. A list of relevant codes is included in the submission form.
- The numerator is defined using ICD-9 diagnosis and procedure codes, HCPCS/CPT procedure codes, and vital status data from the Medicare Enrollment Database. A list of complications included in the numerator within 30 days and 90 days is included in the submission form.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- The developer notes that the ACC's National Cardiovascular Data Registry (NCDR), from which measure data are also derived, includes two programs to ensure data quality. The developer states hospitals must achieve >95% completeness of "core field" data elements for warehouse analysis, though descriptive characteristics are not provided. The 2 programs include:
  - The Data Quality Program (DQP), which assesses the completeness and validity of electronic data submitted by participating hospitals; and
  - The Data Audit Program (DAP), which consists of annual on-site chart review and data abstraction.
- With regard to data element reliability, the developers note that the measure has been developed to utilize only claims data elements that have both face validity and reliability, avoiding the use of fields that are thought to be coded inconsistently across hospitals or providers; the developer also cites hospital auditing programs used by CMS to assess overall claims code accuracy. NQF guidance indicates that if data element validity testing has been conducted, additional data element reliability testing is not required.
- The developer also assesses data element reliability by comparing model variable frequencies and odds ratios in two years of data to determine their degree of consistency over time. [*Note: NQF does not typically consider temporal consistency to be a valid method of demonstrating reliability of data elements.*]
  - Data were drawn from the National Cardiovascular Data Registry (NCDR) ICD Registry and from Medicare Part A claims over the time period 2010Q2-2011Q4. The combined two-year sample included a total of 43,711 admissions to 1,279 hospitals.
  - The developer does not provide specific frequencies and odds ratios for each data element in the submission form, but states that overall, risk factor frequencies changed little across years and there were no notable differences in the odds ratios across years of data.
- The developer defines performance score reliability as the degree to which repeated measurements of the same entity agree with each other.
- In line with this thinking, the developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. For testing purposes, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients.
- The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. [*Note: NQF considers this to be an appropriate method of assessing reliability*.]
  - Data were drawn from the National Cardiovascular Data Registry (NCDR) ICD Registry and from Medicare Part A claims over the time period 2010Q2-2011Q4. The combined two-year sample included a total of 43,711 admissions to 1,279 hospitals.
  - The developer randomly split this sample into two groups, leaving 21,856 admissions to 1,254 hospitals in one randomly selected sample and 21,855 admissions to 1,246 hospitals in the remaining sample.
  - The developer calculated the measure for each hospital in the first sample, and then repeated the calculation using the second sample; thus, each hospital is measured twice, but each measurement is

made using an entirely distinct set of patients.
    - A table is provided showing the distribution of risk-standardized complication rates (RSCRs) within these randomly-split samples.
    - The developer suggests that, to the extent that the calculated measures of these two subsets agree, it shows that the measure is assessing an attribute of the hospital, not of the patients. Agreement was calculated using an intra-class correlation coefficient (ICC).
        - However, the developer seems to suggest that hospitals with fewer than 25 cases were excluded from the sample for purposes of calculating agreement; after excluding these hospitals, the first sample contained 297 hospitals and the second sample contained 298 hospitals. [*Note: The developer does not specify how many patients were included in the testing sample after excluding hospitals with fewer than 25 cases; it is also unclear whether the measure itself excludes hospitals with fewer than 25 cases*.]
    - The developer reports that the agreement between the two RSCRs for each hospital was **0.1494**; the developer notes that, according to the conventional interpretation, this is considered "slight" agreement.
    - The developer also notes that the ICC is based on a split sample of 2 years of data, resulting in a volume of patients in each sample equivalent to only 1 year of data, whereas the measure is likely to be reported with a full two years of data.
- The developer's interpretation of reliability testing results is that the stability over time of the risk factor frequencies and odds ratios indicate that the underlying data elements are reliable, and that the ICC score demonstrates fair agreement across samples using a "strict" approach to assessment that would likely improve with greater sample size.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- For each measured entity, this measure calculates a risk-standardized rate of complications within 30 or 90 days (depending on the complication) following initial ICD implantation. The measure treats complications as a dichotomous (yes/no) variable; the developer notes that their interest is in whether or not a complication has occurred and not how many complications occurred in each hospital.
- As a rationale for measuring this outcome, the developer suggests that complications in the ICD patient population can be reduced through comprehensive, personalized risk assessment and competency of the physician and/or hospital treating the patient.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- With regard to data element validity, the developer refers to programs established within ACC's National Cardiovascular Data Registry (NCDR) that are intended to ensure data quality:
    - The Data Quality Program (DQP), which assesses the completeness and validity of electronic data submitted by participating hospitals; and
    - The Data Audit Program (DAP), which consists of annual on-site chart review and data abstraction.
- In addition, the developer conducted a chart validation study to determine whether ICD-9-CM diagnosis

and procedure codes reported on Medicare claims and used in the measure specifications accurately identify patients experiencing ICD complications within 30 or 90 days of ICD implantation as reported in the medical charts.

- The developer notes that achieving a "substantial" degree of agreement (based on the conventional interpretation of a Kappa coefficient) would have required a sample size of approximately 1,000 charts. However, this was beyond the developer's budgetary restraints; for the final analysis, a sample of 411 medical records from eight hospitals was reviewed.
    - The developers report that the study found an overall agreement between chart and claims (based on paired ratings) of **91.5%**, with a Kappa coefficient of **0.83** (0.7865 – 0.8907) which the developer notes is in the "almost perfect" range, according to the conventional interpretation. The developer states the data comparison included all applicable claims and clinical information, though they do not state all critical data elements (numerator, denominator and exclusions) were tested separately.
- The developer also refers to analyses conducted as part of their risk model diagnostics.
    - For results, see the section on Risk Adjustment below.
- Interpreting their validity testing results, the developer suggests that the audits conducted by the ACC support the overall validity of the data elements included in this measure, and that the data elements used for risk adjustment were consistently found for all patients and were accurately extracted from the medical record.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- This measure has the following exclusions:
    - Hospital stays in which the patient had an ICD implanted prior to the index hospital stay
    - Hospital stays in which the patient had a previous pacemaker placement prior to the index hospital stay
    - Patient admissions in which the patient is not enrolled in Medicare FFS at the time of the ICD procedure
    - Patients who cannot be tracked for 90 days following discharge
    - Not the first claim in the same claim bundle
- The developer provides a brief rationale for each of these exclusions in the submission form as well as details on how the exclusions are identified.
- To ascertain the impact of exclusions on the cohort, the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion.
- A table showing exclusions from the target cohort for the combined 2010-2011 study sample is included in the submission form; the table appears to show the number of patients excluded by each criterion, and the resulting number of patients and hospitals remaining in the analysis after each exclusion criterion is applied sequentially.
- The developer states that the majority of exclusions are necessary to 1) link registry and administrative data (e.g. excluding patient not enrolled in Medicare FFS) and 2) identify patients eligible for complication (e.g. excluding patients who died before discharge).
- The developer suggests that as a result, these exclusions are not discretionary and do not require further testing.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the*

data collection burden)?

2b4. Risk adjustment:

Clinical risk-adjustment:

- The developer states that their risk-adjustment approach uses a parsimonious model that included key variables previously shown to be associated with complications following ICD implantation.
- The developer considered 15 variables for inclusion in their risk-adjustment model, using logistic regression with stepwise selection (entry p<0.05; retention with p<0.01) for variable selection, meaning variables were retained in the model if they met a certain threshold of statistical significance within the model.
- The developer notes that the variables included in the risk model were fully harmonized with the NCDR's existing risk model for in-hospital adverse events; several variables were not clinically significantly associated with risk of complications at 30/90 days, but the developer elected to retain them in the model for consistency with the NCDR's existing model.

Adjustment for Sociodemographic Status (SDS) Factors:

- The developer states that SDS factors were carefully considered by the stewards of this measure, and upon clinical review from the expert panel, found to be less associated with complications.
- The developer argues that routine inclusion of SDS variables in risk models has the potential to explain away meaningful and actionable differences in hospital performance, that analyses have shown that many hospitals caring for a higher proportion of disadvantaged patients still perform well on the measure, and that literature shows that the inclusion of SDS does not consistently meaningfully improve the discriminatory capacity of risk adjustment models.
- The developer notes that the Agency for Healthcare Research and Quality (AHRQ)-validated SES index score provides an analytic approach to accounting for SDS status.
- The developer used this index score to show the distribution of risk standardized complication rates (RSCR) by SES status; the results of this analysis are provided in the form of a box-and-whiskers chart.
- The developer states that the results show little appreciable variation in distribution of risk standardized complication rates, and demonstrate that adjustment for socioeconomic status does not have a statistically meaningful impact on the complication measure results.
- The developer states this as the conceptual reason and justification for **not including SDS status** as a variable within the model.

Risk Model Diagnostics:

- To assess the overall performance of their risk-adjustment model, the developers computed several summary statistics, including:
    - Area under the receiver operating characteristic (ROC) curve (also known as a c-statistic, which measures the probability that the model's prediction of the outcome is better than chance)
        - A c-statistic of 0.64 means that for 64% of all possible pairs of patients—one who suffered a complication and one who didn't—the model correctly assigned a higher probability to those who had a complication. Generally, a c-statistic of at least 0.70 is considered acceptable.
    - Predictive ability (the model's ability to distinguish high-risk subjects from low-risk subjects)
        - The developer notes that a wide range between the lowest decile and highest decile is an indication that the model has good predictive ability.
    - Over-fitting indices (to ensure that the model is not only describing the relationship between predictive variables and outcome in the development dataset but also providing valid predictions in new patients)
        - Presented as $(\gamma 0, \gamma 1)$ – the developer suggests that if the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from 1, there is potential evidence of over-fitting. A calibration value close to 0 at one end and close to 1 on the other end indicates good calibration of the model.
- The results of these analyses are as follows:
    - Derivation sample:
        - C-statistic = **0.640**

        - Predictive ability (lowest decile %, highest decile %): **4.05%, 25.08%**
    - Validation sample:
        - C-statistic = **0.642**
        - Predictive ability (lowest decile %, highest decile %): **3.80%, 23.80%**
    - Over-fitting indices [model calibration]: **(0.03, 1.02)**
- The developer states that the C-statistics of 0.640 and 0.642 indicate "fair" model discrimination in the derivation and validation cohorts, suggesting that complications, as opposed to other outcomes such as mortality, consistently have a lower c-statistic, likely because complications are less determined by patient comorbidities and more by health system factors.
- The developer notes that the model showed a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk patients from low-risk patients.
- The developer suggests that, interpreted together, their diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

Questions for the Committee:

- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care?*
- *Does the Standing Committee agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference:

- The developer notes that for public reporting purposes, an interval estimate for each hospital's risk-standardized complication rate is estimated to characterize the amount of uncertainty associated with the rate; these interval estimates are then compared to the national crude rate for the outcome, with hospitals being categorized as "better than," "worse than," or "no different than" the U.S. national rate.
- To assess variation in RSCRs among hospitals, the developer examined the distribution of hospital RSCRs and plotted these values in the form of a histogram.
- The developer states that recent analyses of Medicare FFS data show variation in RSCRs among hospitals, noting that based on data from 2010Q2-2011Q4, the mean hospital RSCR was 6.8%, with a range of 5.42% to 9.52% and an interquartile range of 6.55% to 6.99%.
- The developer states that the variation in rates suggests there are clinically meaningful differences across hospitals for 30/90 day risk-standardized complications after ICD insertion.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- To identify the extent and distribution of missing data, the developer examined rates of missing data for all candidate variables and examined histograms of the frequency of missing data by hospital.
- The developer reports that overall, the percentage of missing values for all categorical variables was very small (<1%) and were imputed to specific categories based on the developer's previous experience.
- The developer suggests that model performance was comparable when they included or excluded cases with missing data.

2d.Composite measure: construction

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

construct.

- The developer states that the empirical analysis demonstrating the individual component measures fit the overall quality construct is currently underway. The developer reports that this data can be provided at the next maintenance review, once testing is completed.
- The empirical analysis will focus on construct validation which will test the hypothesis on the theory of the construct that following these processes for patients with ICD implantations lead to better outcomes.
- The developer believes they have achieved parsimony by including as few elements as possible without impacting the psychometric properties of the measure.

Questions for the Committee:

- *Do the component measures fit the quality construct?*
- *Are the objectives of parsimony and simplicity achieved while supporting the quality construct?*

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- The elements are clearly defined. They all come from the NCDR.
- Use ICD-9 diagnosis and procedure codes to identify values in both the numerator and denominator. Will need to address how to use ICD-10 codes moving forward. Match Medicare FFS patients with NCDR ICD registry. The only concern is it looks at patients who are Medicare FFS patients. Utilize claims data elements with face reliability.

2a2.: ***Committee's Comments on Reliability-Testing:***

- Reliability was tested with a development sample and a test sample. Numerically, the results were very similar. The developer reports that the agreement between the two RSCRs for each hospital was 0.1494. I am surprised that this statistic is so small given the similarity of the distributions of the two samples.
- Agreement rate for scores for each hospital was 0.1494 and this is determined as slight agreement.

2b1.: ***Committee's Comments on Validity-Specifications***:

- I didn't see any threats to validity. This is not a PRO
- The outcome is measured as a dichotmous variable (Y/N) rather than how many complications which they authors conclude that it makes this measure valid as it allows for complication rate comparisons across hospitals.

2b2.: ***Committee's Comments on Validity-Testing:***

- The data are drawn from the NCDR which has both a Data Quality Program (DQP) and a Data Audit Program (DAP). CMS audits the codes that a hospital submits. In a sample of 411 medical records from 8 hospitals, overall agreement between chart and claims was 0.83. "nearly perfect"
- Yes, validity was tested and showed an overall agreement between chart and claims (based on paired ratings) of 91.5%, with a Kappa coefficient of 0.83 (0.7865 – 0.8907) which the developer notes is in the "almost perfect" range.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Exclusions are appropriate as is the risk adjustment model.
- The problem that the developer is having in using the data is that they have been told by ResDAT that they can't use ResDAT for quality assessment. The ACC is currently in the process of establishing an agreement that will give them access to claims data--access that is necessary if this measure is to be implemented.
- The way in which missing data and empty fields will be handled is clearly specified and is appropriate.
- The measure steward has not analyzed the performance of every component (because they didn't realize that this was a composite measure until a few days ago), but because each of the components is based on a claim, I don't see that this is a significant issue.
- One potential threat to validity is that patients are excluded if they do not have 90 day follow-up which could underrepresent the rate of complications. Also, patients are excluded if they are not Medicare FFS patients based on the data linkage and tracking of outcomes
- Considered 15 variables for a risk-adjustment model using logistic regression with stepwise selection for each

variable (retention with [< 0.01).
- The developer states that the C-statistics of 0.640 and 0.642 indicate "fair" model discrimination in the derivation and validation cohorts
- Recent analyses of Medicare FFS data show variation in RSCRs among hospitals, noting that based on data from 2010Q2-2011Q4, the mean hospital RSCR was 6.8%, with a range of 5.42% to 9.52% and an interquartile range of 6.55% to 6.99%.
- Missing data is less than 1% of patients analyzed

2d.: ***Committee's Comments on Composite Performance Measure:***
- All of the analysis is, in fact, directed at the composite. The analysis demonstrates that the measure is reliable and valid.
- An empirical analysis demonstrating the individual component measures fit the overall quality construct is currently underway and will be available at the live meeting per the submission information.

Criterion 3. Feasibility

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

- The data sources for this measure are ICD-9 diagnosis and procedure codes, HCPCS/CPT procedure codes, and vital status data from the Medicare Enrollment Database.
- The developer states that all data elements are in defined fields in electronic clinical data.
- The measure is specified to collect data through the ACC's National Cardiovascular Data Registry (NCDR) ICD Registry, which requires fees for participation; however, the developer notes that the ACCF also allows for licensing of the measure specifications outside of the Registry.
- The developer also suggests that centers already have to participate in this specific registry for reimbursement purposes, so that currently almost all hospitals that implant ICDs in Medicare populations already participate in the Registry.

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***
- All of the data elements are generated either from the NCDR or CMS claims. The measure is only for patients who were enrolled in Medicare before the beginning of the episode. The measure is not feasible for the commercial population because of impracticality of following patients after discharge from hospital.
- Data is available through the Medicare Enrollment Database and thus it is feasible to extract this data. In addition data can be extracted from the NCDR ICD registry.

Criterion 4: Usability and Use

4. Usability and 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.

- This measure is not currently in use, but the developer states their commitment to implementing the measure, and notes that ACC is currently in the process of applying to be a Qualified Entity for Medicare data reporting purposes.
- With regard to potential unintended consequences, the developer notes that publicly reporting hospital risk-standardized ICD complication rates requires that the data submitted by hospitals be complete, consistent, and

accurate.

- The developer refers to methods used to ensure data quality as part of the NCDR's existing Data Quality Program (DQP) to address these issues.

Questions for the Committee (as appropriate) *:*

- *Is the measure publicly reported?*
- *For maintenance measures – is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- The measure is not being used at this time because the steward has not been able to gain access to Medicare data. The steward is working with CMS to gain access to the data.
- Planned use includes Public Reporting; Quality Improvement with Benchmarking (external benchmarking to multiple organizations); and Quality Improvement (Internal to the specific organization)
- The measure is not currently in use and discusses using NCDR data quality program requirements in order to maintain reliability and usability of the data.

Criterion 5: Related and Competing Measures

- List any related or competing measures based on harmonization protocol.
- Summarize any harmonization efforts, i.e., responses from the developers regarding harmonization.
- Briefly summarize next steps according to protocol

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0694

Measure Title: Hospital Risk-Standardized Complication Rate following Implantation of Implantable Cardioverter-Defibrillator (ICD)

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

Date of Submission: Click here to enter a date

Instructions

- *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.*

- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☒ Health outcome: Avoidance of complications

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

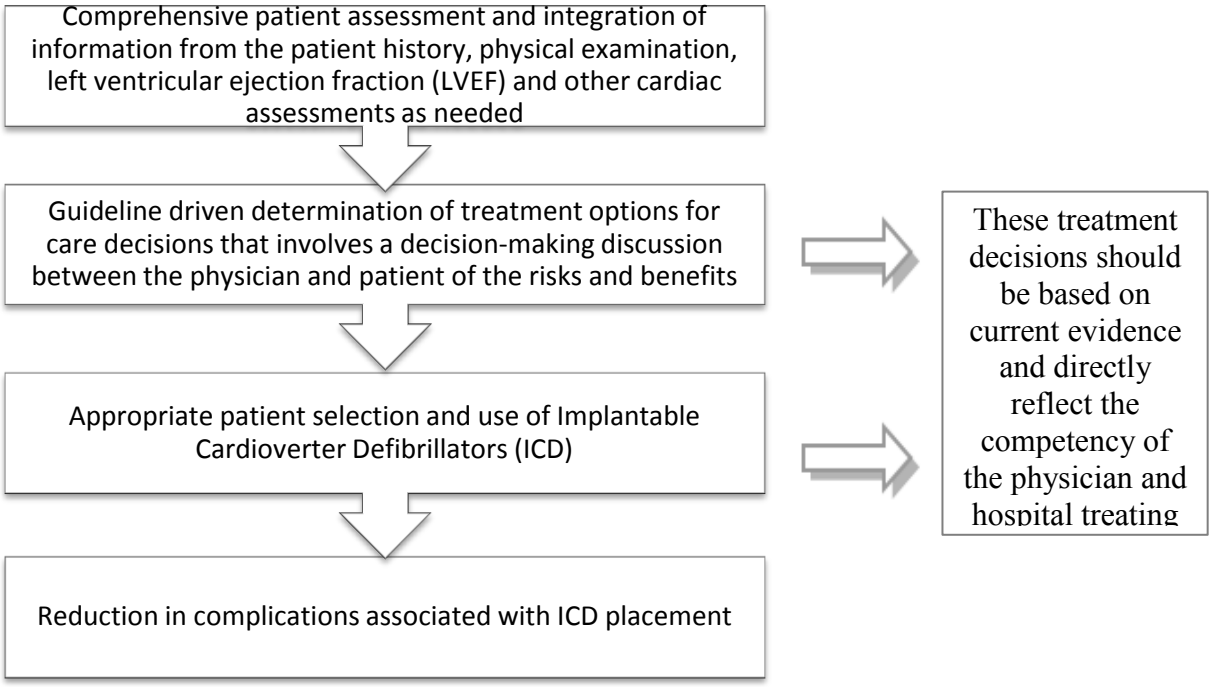
☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

A complication following placement of an Implantable Cardioverter Defibrillator (ICD) is an undesirable outcome. Comprehensive, personalized risk assessment and competency of the physician and hospital treating the patient can lead to decreased complications in the ICD patient population.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Complications following insertion of Implantable Cardioverter Defibrillators (ICD) are an important patient outcome (Al-Khatib 2005, 2008; Curtis 2009, Peterson, 2013) that reflects the quality of care delivered to patients.

ICDs are expensive and are utilized in patients with high cost conditions such as coronary artery disease or heart failure. Reynolds, et al (2006) found that just over 10% of all patients with an ICD placed had a complication deemed attributable the procedure. The cost to treat these unexpected complications was more than $7,000 for each patient and frequently extended a patient's hospital stay.

The risk of adverse outcomes following ICD implantation varies markedly by the experience and training of the implanting physician, the device implanted, and the characteristics of the facility in which the procedure is performed (Curtis, 2009). These structures and processes of care that prevent these costly and undesirable

complications are difficult to measure in a way that is reliable, valid and meaningful to providers and patients. It is important that ICDs are provided to those patients for whom it is deemed appropriate based on guideline-based assessments and evaluations and hospitals ensure provision of the highest quality of care. Reporting the rate of procedure-related complications provides relevant information on whether these characteristics were achieved.

Al-Khatib SM, Greiner MA, Peterson ED, Hernandez AF, Schulman KA, Curtis LH. Patient and Implanting Physician Factors Associated With Mortality and Complications After Implantable Cardioverter-Defibrillator Implantation, 2002-2005. Circ Arrhythmia Electrophysiol. 2008;1:240-249. doi: 10.1161/CIRCEP.108.777888

Al-Khatib, SM, Lucas LF, Jollis JG, Malenka DJ, Wennberg DE. The relation between patients' outcomes and the volume of cardioverter-defibrillator implantation procedures performed by physicians treating Medicare beneficiaries.[see comment][erratum appears in J Am Coll Cardiol. 2005;46:1964]. Journal of the American College of Cardiology, 2005;46: p 1536-40.

Curtis JP, Luebbert JJ, Wang Y; et al. Association of physician certification and outcomes among patients receiving an implantable cardioverter-defibrillator. JAMA. 2009;301(16):1661-1670.

Reynolds MR, et al. Complications among Medicare beneficiaries, receiving implantable cardioverter-defibrillators. J Am Coll Cardiol. 2006; 47:2493-2497.

Peterson PE, et al. Association of single- vs dual-chamber ICDs with mortality, readmissions and complications among patients receiving an ICD for primary prevention. JAMA. 2013;309:2025-2034.

Additional relevant articles:

Pokorney SD, et al. Primary prevention implantable cardioverter-defibrillators in older racial and ethnic minority patients. Circ Arrhythm Electrophysiol. 2015 Feb;8(1):145-51

Dodson, JA, et al. 2014. Developing a Risk Model for in-Hospital Adverse Events following ICD Implantation: A Report from the NCDR® Registry. Journal of the American College of Cardiology, 63(8), 788–796. doi:10.1016/j.jacc.2013.09.079

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0694_ACC_Hospital_Complication_ICD_evidence_submission_061715_submitted.docx

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

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Reported complication rates following ICD implantation vary from 4% to 30%, depending on how complications are defined and the period of assessment. In the ACC's NCDR ICD Registry, the incidence of in-hospital complications is approximately 4%. However, complications such as device infection, malfunction, or cardiac tamponade may only become evident following hospital discharge. Al-Khatib et al. (2008) found overall rates of complication within 90 days of ICD implantation ranged from 18.8% in 2002 to 14.2% in 2005. Additional updated evidence and peer-reviewed publications can be found in the evidence supplement.

Preliminary analyses confirm that serious complications, including death, can occur after ICD implantation, and that there is substantial variation in complication rates across hospitals. As the foundation of this measure, we conducted analyses to determine unadjusted ICD-related complication rates in Medicare inpatient claims data for 2007, which included 67,532 ICD admissions for 67,080 patients at 1,792 hospitals. Administrative codes identifying ICD-related complications were identified through review of the literature and subsequently refined in conjunction with input from topic experts to capture the most significant complications. Complications were identified from CMS claims data using ICD-9-CM diagnosis and procedure codes or mortality within specified timeframe (30 days or 90 days following implantation depending on the specific complication). In these preliminary analyses, complications were seen in 5.7% of ICD admissions (3,818 complications).

In addition to clinically important rates of complications following ICD implantation, there is substantial variability in rates across hospitals. The median complication rate following ICD implantation ranges from 0% to 17.8% across deciles of hospitals grouped by their all-cause complication rate.

These findings suggest that these complications are reasonably attributable to the ICD, potentially preventable, and thus, actionable.

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.**
Al-Khatib SM, Greiner MA, Peterson ED, Hernandez AF, Schulman KA, Curtis LH. Patient and Implanting Physician Factors Associated With Mortality and Complications After Implantable Cardioverter-Defibrillator Implantation, 2002-2005. Circ Arrhythmia Electrophysiol. 2008;1:240-249.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
We examined health disparities associated with this measure by race/ethnicity and by socioeconomic status (SES). For both sets of analyses, the distribution of race and SES were stratified by quintiles. SES status (SES score) was determined from AHRQ SES Index

data, and the percentage of African American patients was derived from CMS FFS patients for any condition during 2010 and 2011.

Dates of Data: 2010Q2 – 2011Q4
Number of Measured Entities (Hospitals): 1279
Number of Patients: 43711

A box and whisker plots depiction can be found in the testing supplement.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Patient/societal consequences of poor quality, Frequently performed procedure, High resource use
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Over the past two decades, clinical trials have demonstrated that ICDs reduce the risk of sudden cardiac death for select high risk patients. As a result of these trials, there is a large increase in the number of patients undergoing ICD implantation, with an estimated increase in the number of inpatient implantations from 5,600 in 1990 to 108,680 by 2005 (Brown, Croft et al. 2008). Although ICD therapy can improve the survival of appropriately selected patients, device implantation carries a low but unavoidable risk of significant complications, which are associated with increased cost, length of stay, and higher risk of mortality (Al-Khatib, Greiner et al. 2008).

ICD implantation is an expensive procedure performed on patients with advanced cardiovascular disease and, often, significant comorbidities. Despite improvements in technology and increasing experience with device implantation, the procedure carries a significant risk of complications (Hammill, Curtis, 2008).

--Roughly 150,000 ICDs are implanted each year and approximately two thirds of implantations are performed on Medicare patients

--Direct total medical cost per device (2005) (Sanders et al, 2005) is $68,000-$100,000. The total costs to payers ranges from $10-$15 billion, of which $7-$10 billion is fee-for-service Medicare

-- Costly complications are common with 11% of Medicare patients having early complications

--In one study (Reynolds et al, 2006) complications increased length of stay 1-10 days and raised costs $5,000 – 20,000 (mean $7,251), adding roughly $80 million in Medicare costs

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Hammill S and Curtis J. Publicly Reporting Implantable Cardioverter Defibrillator Outcomes – Grading the Report Card. Circ Arrhythmia Electrophysiol. 2008;1:235-237).

Sanders GD, Hlatky MA, Owens DK. Cost-Effectiveness of Implantable Cardioverter-Defibrillators. N Engl J M. 2005;353;1471-1480.

Reynolds, M.R., et al., The frequency and incremental cost of major complications among medicare beneficiaries receiving implantable cardioverter-defibrillators. Journal of the American College of Cardiology, 2006. 47(12): p. 2493-7.

Brown, D.W., Croft, J.B., et al. (2008). "Trends in Hospitalizations for the Implantation of Cardioverter-Defibrillators in the United

States, 1990-2005." American Journal of Cardiology 101 (12): 1753-1755.

Al-Khatib SM, Greiner MA, Peterson ED, Hernandez AF, Schulman KA, Curtis LH. Patient and Implanting Physician Factors Associated With Mortality and Complications After Implantable Cardioverter-Defibrillator Implantation, 2002-2005. Circ Arrhythmia Electrophysiol. 2008;1:240-249.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

1d. Composite Quality Construct and Rationale

1d.1. **A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
    - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient); or
    - any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient).

1d.1. Please identify the composite measure construction: any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

1d.2. **Describe the quality construct, including:**

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

Research strongly suggests that the complications--as identified in this application--are reasonably attributable to the ICD, potentially preventable, and thus actionable.

1) In defining the complications, we sought clinically sensible definitions which were, to the extent possible, likely attributable to ICD implantation. In consultation with an expert panel, it was agreed that restricting outcomes to complications requiring an intervention would enhance measure acceptance as these complications represent the most clinically significant adverse events.
The original list of complications routinely captured in administrative data included:

1. Pneumothorax
2. Hematoma
3. Tamponade
4. Mechanical complications
5. Pulmonary embolism
6. Infection
7. Other cardiac complication
8. Acute renal failure requiring hemodialysis
9. Death

For several of the complications identified in this measure, we chose to refine the definitions following ICD implantation (as discussed in the literature) to include associated interventions. The resulting modified definitions represent the most clinically significant complications. For example, the claims code used to identify "mechanical complications" in the claims data is broad, including both lead dislodgement requiring open revision as well as minor lead abnormalities that can be addressed simply by reprogramming the device. As such, restricting to mechanical complications with a system revision narrows the focus of the measure on complications of device implantation that require intervention. Similarly, restricting "pneumothorax or hemothorax" and "hematoma" complications to those requiring an intervention was deemed important because those events vary widely in identification, clinical severity, and recommended treatment. "Pulmonary embolism" and "acute renal failure requiring hemodialysis" were dropped from the list of complications as the observed rates were low (less than 0.15%) and they were deemed less clearly attributable to the ICD implantation itself. "Other cardiac complication" was also dropped because it was deemed to be too broad for

this measure. Finally, we added two additional complications to the original list: additional ICDs implanted within 90 days of the index procedure and death within 30 days of the index procedure. In both cases, the event would be an unplanned, adverse event. The final list of complications is as follows:

1. Pneumothorax or hemothorax, with chest tube
2. Hematoma with blood transfusion or evacuation
3. Cardiac tamponade or pericardiocentesis
4. Mechanical complications requiring a system revision
5. Infection that is device related
6. Second ICD within 90 days of the index procedure
7. Death

(2) In consultation with an expert panel, we chose a hybrid complication-specific approach for the outcome time period. Review of preliminary analyses revealed that most complications occur within the initial 15 days following implantation, and qualitatively plateaued between 30 and 45 days following ICD implantation. Accordingly, we initially considered a 30-day time period for follow-up. However, feedback from topic experts suggested that using a single period of assessment for such a wide range of outcomes may not be the optimal approach. For example, device related infections may not become apparent for weeks or months following implantation, suggesting a 90 day time period would be best for this outcome. In contrast, however, hematomas due to the procedure would most likely be recognized and treated within 30 days of implantation, and hematomas identified after that point are more likely due to other procedures (e.g., cardiac catheterization). Given these considerations, we adopted timeframes specific to each complication. The timeframes are as follows:

30-day timeframe
- Pneumothorax or hemothorax, plus chest tube
- Hematoma plus blood transfusion or evacuation
- Cardiac tamponade or pericardiocentesis
- Death

90-day timeframe
- Mechanical complications requiring system revision
- Device related infections
- Additional ICD implantations

Citations

Al-Khatib SM, Greiner MA, Peterson ED, Hernandez AF, Schulman KA, Curtis LH. Patient and Implanting Physician Factors Associated With Mortality and Complications After Implantable Cardioverter-Defibrillator Implantation, 2002-2005. Circ Arrhythmia Electrophysiol. 2008;1:240-249.

Al-Khatib, S.M., et al., The relation between patients´ outcomes and the volume of cardioverter-defibrillator implantation procedures performed by physicians treating Medicare beneficiaries.[see comment][erratum appears in J Am Coll Cardiol. 2005 Nov 15;46(10):1964]. Journal of the American College of Cardiology, 2005. 46(8): p 1536-40.

Curtis JP, Luebbert JJ, Wang Y; et al. Association of physician certification and outcomes among patients receiving an implantable cardioverter-defibrillator. JAMA. 2009;301(16):1661-1670.

1d.3. Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.

Composite performance measures have a variety of uses.

Data reduction. A large and growing array of individual indicators makes it possible for users to become overloaded with data. A composite measure reduces the information burden by distilling the available indicators into a simple summary.

Scope expansion. The information in a composite measure is highly condensed, making it feasible to track a broader range of metrics than would be possible otherwise. Composite measures have been described as a tool for making provider assessments more comprehensive

Provider performance valuation. Performance indicators are used for various decisions about providers, including the allocation of pay-for-performance incentives, designation of preferred provider status, and assignment of letter grades and star rating categories. If a decision is to be based on multiple indicators instead of a single indicator, a method of translating several variables into a single decision is needed. Composite measures serve this function by assigning providers to 1 position on a scale of better-to-worse performance.

Each complication, when viewed in isolation, offers a narrow perspective of the performance of the given hospital (s). This measure attempts to focus on various key clinical negative outcomes where the locus of control is within the given operator or site.

1d.4. Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.

Each of the components of this measure address key patient outcomes. Many factors go into the care of patients. This measure was developed to identify institutions whose performance is better or worse than would be expected based on their patient case mix, regardless of complications. The clinical subject matter experts believed that no complication should be treated less "severe". The measure specifications reflect this notion.

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 subcriteria 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 : Ischemic Heart Disease, Coronary Artery Disease

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

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

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** icd_v2_datadictionary_codersdictionary_2-1-635699788053782318.pdf

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Since this measure was last endorsed on 01/2011, ACC has updated this measure to a more parsimonious version. The original 13 variables have been reduced to 9 variables. Over the next years, it is a priority of the ACC to focus the development of measures that require less variables while ensuring statistical reliability and validity properties are not reduced. One additional benefit for sites is that the burden of data capture may be less resource intensive and could result in less “missing” data.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The outcome for this measure is one or more complications within 30 or 90 days (depending on the complication) following initial ICD implantation. The measure treats complications as a dichotomous (yes/no) variable; we are interested in whether or not a complication has occurred and not how many complications occurred in each hospital.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The time period for aggregating data for public reporting has not been determined. Other outcome measures have been calculated using 2 or 3 years of data, and it is likely that this will provide a sufficient volume of cases needed to calculate hospital RSCR.

S.6. **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, 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.
Complications are identified using International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) diagnosis and procedure codes or Healthcare Common Procedure Coding System/Current Procedural Terminology (HCPCS/CPT) procedure codes as well as the Medicare Enrollment Database (vital status) as indicated below. This approach was developed by a CMS Technical Expert Panel of clinicians and methodologists who were charged with identifying a comprehensive claims-based approach to identifying serious procedural complications:

Complications identified within 30 days of device implant

(1) Pneumothorax or hemothorax plus a chest tube
Definition: (a) Pneumothorax / hemothorax: 512.0, 512.1x, 512.8, or 511.8x (diagnosis code) AND
(b) Chest tube: 34.04, 34.05, 34.06, or 34.09 (procedure code)
(2) Hematoma plus a blood transfusion or evacuation
Definition: (a) Hematoma: 998.1x (diagnosis code) AND
(b) Blood transfusion: 518.7x, 287.4x, V59.01, V58.2x (diagnosis code), or
99.00, 99.03, 99.04 (procedure code) OR
Evacuation: 34.04, 34.09 (procedure code)
(3) Cardiac tamponade or pericardiocentesis
Definition: (a) Cardiac tamponade: 420.xx, 423.0x, 423.3x, 423.9x (diagnosis code) OR
37.0, 37.12 (procedure code)
(4) Death
Source: Medicare enrollment database

Complications identified within 90 days of device implant

(5) Mechanical complications requiring a system revision
Definition: (a) Mechanical complications with system revision: 996.0x, 996.72 (diagnosis code) AND
(b) System revision: 37.75, 37.77, 37.79, 37.97, 37.94, 37.99, 39.94, or
00.52(procedure code)
(6) Device related infection
Definition: (a) Infection: 996.61 (diagnosis code)
(7) Additional ICD implantation
Definition: (a) Inpatient or outpatient ICD implantation: 00.50, 00.51, 00.52, 00.53,00.54, or 37.94 (procedure codes) OR
(b) Outpatient ICD implantation: 33216, 33217, 33218, 33220, 33223, 33230, 33231, 33240, 33241, or 33249, 33262, 33263, 33264 (CPT procedure codes)

We used the General Equivalence Mapping (GEM) crosswalk between ICD-9-CM and ICD-10-CM/PCS to create specifications for the ICD complication measure in ICD-10-CM/PCS. Additionally, our process for mapping procedural codes in the measures to ICD-10 included detailed clinical review, including manual review of related ICD-10 codes to determine that all appropriate codes were included, rather than relying exclusively on the GEM. See appendix A.1. supplemental files.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The target population for this measure includes inpatient and outpatient hospital stays with ICD implants for patients at least 65 years of age who have matching information in the National Cardiovascular Disease Registry (NCDR) ICD Registry. The time window can be specified from one to three years. This measure was developed with Medicare claims and CathPCI Registry data from one calendar year (2007).

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
We use this field to define the measure cohort, defined by ICD-9 procedures codes from inpatient claims and HCPCS/CPT procedure

codes from outpatient claims as outlined below:
ICD-9 codes
00.50 Implantation of cardiac resynchronization pacemaker without mention of defibrillation, total system (crt-p)
00.51 Implantation of cardiac resynchronization defibrillator, total system (crt-d)
00.52 Implantation or replacement of transvenous lead (electrode) into left ventricular coronary venous system
00.53 Implantation or replacement of cardiac resynchronization pacemaker pulse generator only (crt-p)
00.54 Implantation or replacement of cardiac resynchronization defibrillator pulse generator device only (crt-d)
37.94 Implantation or replacement of automatic cardioverter/defibrillator, total system (aicd)
CPT codes33216 Insertion, single chamber transvenous electrode ICD
33217 Insertion, dual chamber transvenous electrode ICD
33218 Repair, single chamber transvenous electrode ICD
33220 Repair, dual chamber transvenous electrode ICD
33223 Pocket revision ICD
33230 Initial pulse generator insertion only with existing dual leads
33231 Initial pulse generator insertion only with existing multiple leads 33240 Insertion of single or dual chamber ICD pulse generator
33241 Removal of single or dual chamber ICD pulse generator
33249 Insertion or repositioning of electrode lead(s) for single or dual chamber pacing ICD and insertion of pulse generator
33262 Removal pulse generator with replacement pulse generator only single lead system (transvenous)
33263 Removal pulse generator with replacement pulse generator only dual lead system (transvenous)
33264 Removal pulse generator with replacement pulse generator only multiple lead system (transvenous)

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
(1) Previous ICD placement. Hospital stays in which the patient had an ICD implanted prior to the index hospital stay are excluded.
Rationale: Ideally, the measure would include patients with a prior ICD, as this is a population known to be at high risk of adverse outcomes. However, for these patients it is difficult to distinguish in the administrative data whether adverse events such as infection were present on admission or complications of the second ICD placement. In order to avoid misclassification, we exclude these patients from the measure.
(2) Previous pacemaker placement, Hospital stays in which the patient had a previous pacemaker placement prior to the index hospital stay are excluded.
Rationale: Some complications (infection or mechanical complication) may be related to a pacemaker that was removed prior to placement of an ICD. Ideally, the measure would include patients with a prior pacemaker, as this is a population known to be at higher risk of adverse outcomes. However, for these patients it is difficult to distinguish in the administrative data whether adverse events such as infection were present on admission or complications of the ICD placement. In order to avoid misclassification, we exclude these patients from the measure.
(3) Not Medicare FFS patient on admission. Patient admissions in which the patient is not enrolled in Medicare FFS at the time of the ICD procedure.
Rationale: Outcome data are being derived only for Medicare fee-for-service patients.
(4) Lack 90-day follow-up in Medicare FFS post-discharge. Patients who cannot be tracked for 90 days following discharge are excluded.
Rationale: There will not be adequate follow-up data to assess complications
(5) Not the first claim in the same claim bundle. There are cases when several claims in the same hospital representing a single episode of care exist in the data together. These claims are bundled together and any claim other than the first is excluded.
Rationale: Inclusion of additional claims could lead to double counting of an index ICD procedure.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Denominator exclusions are identified based on variables contained in the Standard Analytic File (SAF) or Enrollment Database (EDB). Of note, a hospital stay may satisfy multiple exclusion criteria.
(1) Previous ICD placement is a flag in the NCDR-ICD registry that indicates whether or not a patient has an ICD present on admission.
(2) Previous pacemaker is a flag in the NCDR-ICD registry that indicates whether or not a patient has a pacemaker present on

admission.
(3) Not Medicare FFS patient on admission is determined by patient enrollment in both Part A and Part B in FFS using CMS' EDB.
(4) Lack 90-day follow-up in Medicare FFS post-discharge is determined by patient enrollment status in both Part A and Part B and in FFS using CMS' EDB; the enrollment indicators must be appropriately marked for any month which falls within 90 days of hospital discharge or enrollment end date (this does not apply for patients who die within 90 days of the index hospital stay).
(5) Not the first claim in the same claim bundle is derived by examining inpatient claims located in the SAF; specifically the fields for admit discharge date and provider ID.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
This measure is not stratified.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Our approach to risk adjustment conforms to the scientific standards for a publicly reported outcome measure as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).
The measure employs a hierarchical logistic regression model to create a hospital-level 30 or 90 day RSCR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, it models the log-odds of hospital complications within 30 or 90 days of discharge using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of complication at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.
Candidate and Final Risk-adjustment Variables: Candidate variables were patient-level risk-adjustors that were expected to be predictive of procedural complications, based on empirical analysis, prior literature, and clinical judgment, including age, sex, and indicators of comorbidity and disease severity. For each patient, covariates are obtained from claims records extending 12 months prior to and including the index admission. For the measure currently implemented by CMS, these risk-adjusters are identified using both inpatient and outpatient Medicare FFS claims data.

The model adjusts for case-mix differences based on the clinical status of patients at the time of admission. We use condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in the supplemental materials. In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care and that are only recorded in the index admission.

The 9 variables included in the risk model are listed below.

(1) Sex
Male
Female
(2) Reason for admission
Admitted for procedure
Cardiac heart failure
Other
(3) NYHA class
I/II
III
IV
(4) Prior Coronary Artery Bypass Graft (CABG)

(5) Abnormal conduction
No
Yes-left bundle
Yes-other
(6) ICD type
Single chamber
Dual chamber
CRT-D
(7) Sodium
<135
135-145
>145
(8) Hemoglobin (5 g/Dl)
(9) BUN (10 mg/Dl)

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

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.
Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
See appendix A.1. supplemental files.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
The measure employs a hierarchical logistic regression model to create a hospital-level 30 or 90 day RSCR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, it models the log-odds of hospital complications within 30 or 90 days of discharge using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of complications at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.
The RSCR is calculated as the ratio of the number of "predicted" to the number of "expected" complications, multiplied by the national unadjusted complication rate. For each hospital, the numerator of the ratio ("predicted") is the number of complications within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the denominator ("expected") is the number of complications expected on the basis of the nation's performance with that hospital's case mix. This approach is

analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital's performance given its case mix to an average hospital's performance with the same case mix. Thus, a lower ratio indicates lower-than-expected complications or better quality and a higher ratio indicates higher-than-expected complications or worse quality.
The "predicted" number of complications (the numerator) is calculated by using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of complication. The estimated hospital specific intercept is added coefficients multiplied by the patient characteristics. The results are transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of complications (the denominator) is obtained in the same manner, but a common intercept using all hospitals in our sample is added in place of the hospital specific intercept. The results are transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the years of data in that period.
Reference:
Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
The measure is not based on a sample or survey.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
ACC developed measures generally defaults missing data to "performance not met" However, this proposed measure is not treated like process measures where the measure is attempting to measure the documentation of a given care process. Rather the care that is being measured is—based on a number of variables—if there are certain complications within 30 or 90 days of the implantation. The rates of missing values for different elements in the data vary depending on the specific variable, but are rare (<2%).

Variables for which the missing value defaults to the median of the non-missing values of the corresponding variable are:
BUN
Hemoglobin
Sodium

Variables for which the missing was default to a specific values:

ICD type: single chamber
Reason for Admission defaults to "Admitted for procedure"
NYHA Classification defaults to "NYHA class I or II"
Prior CABG defaults to "no"
Abnormal conduction defaults to "no".

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The datasets used to create the measures are described below.

(1)NCDR ICD Registry data
The National ICD Registry is a cardiovascular data registry which captures detailed information about patients at least 18 years of age undergoing ICD implantation. This includes demographics, comorbid conditions, cardiac status, and laboratory results. As of May 2015, the registry had collected data from 1,786 hospitals in the United States totaling over 1,330,000 implants (NCDR data outcome reports).

The registry, launched on June 30, 2005, was developed through a partnership of the Heart Rhythm Society (HRS) and the American College of Cardiology Foundation (ACCF) in response to CMS' expanded ICD coverage decision for primary prevention ICD therapy. Data included in the registry are collected by hospitals and submitted electronically on a quarterly basis to NCDR. The patient records submitted to the registry focus on acute episodes of care, from admission to discharge. The NCDR does not currently link patient records longitudinally across episodes of care.

The data collection form and the complete list of variables collected and submitted by hospitals can be found at www.ncdr.com. For more information on these data, please see the attached methodology report.

Of note, hospitals are only required to submit data on all primary prevention ICDs implanted in Medicare patients, and, of the 159 data elements collected by the ICD Registry, only 54 are forwarded to CMS by ACC to determine payment eligibility. Nevertheless, the majority of participating hospitals have opted to participate fully in the quality improvement aspect of the registry, and submit all data elements on all patients undergoing ICD implantation.

(2)Medicare Data
The model was developed in a population of Medicare fee-for-service beneficiaries but can be expanded to all ICD patients at least 65 years of age. We used the administrative claims data to identify complications.

(a) Part A inpatient and outpatient data: Part A data refers to claims paid for Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, and hospice care. For this measure, we used Part A data to identify ICDs implanted for admitted and non-admitted patients (i.e. hospital patients with observation status). For model development, we used 2007 Medicare Part A data to match patient stays associated with an ICD with comparable data from the NCDR ICD Registry.

(b) Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This dataset was used to obtain information on several inclusion/exclusion indicators, such as Medicare status on admission, and provided the ability to retrieve 90 days follow-up, linking patient Health Insurance Claim (HIC) number to the Part A data. These data have previously been shown to accurately reflect patient vital status (Fleming Fisher et al. 1992).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Urgent Care, Hospital/Acute Care Facility
If other:

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0694_Complications_testing__061715_submitted-635705702135277500.docx

Composite Measure Title: 0694

Measure Title: Hospital-level Risk-Standardized Complication Rate following Implantation of Implantable Cardioverter-Defibrillator

Date of Submission: Click here to enter a date

Composite Construction:

☐Two or more individual performance measure scores combined into one score

☐ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

☐ Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient

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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

Notes

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

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

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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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 <u>all</u> 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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☐ administrative claims
☒ clinical database/registry	☐ clinical database/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*).

We propose to use a clinical registry, the National Cardiovascular Data Registry (NCDR) ICD Registry. This is a national quality improvement registry that is currently participated in >1,600 US hospitals. Some states and healthcare systems mandate participation. Rigorous quality standards are applied to the data and both quarterly and ad hoc performance reports are generated for participating centers to track and improve their performance.

The measure links clinical data from NCDR to Medicare claims data to ascertain complications.

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

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

For this measure, hospitals are the measured entities. All non-federal, acute inpatient US hospitals (including territories) that participate in the American College of Cardiology (ACC) NCDR's ICD Registry and care for Medicare Fee-for-Service (FFS) beneficiaries who are 65 years of age or older are included. The number of measured entities (hospitals) varies by testing type; see Section 1.7 for details.

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

Table 1. Patient Characteristics, 30/90 Day Complications

Description	**Total**	
	#	**%**
ALL	43711	100.00
Admission Characteristics		
Female	12123	27.73
Hospital Reason		
Admitted for this Procedure	26658	60.99
Hospitalized-Cardiac heart failure	5173	11.83
Hospitalized-Other	11880	27.18
History and Risk Factors		
NYHA Class - Current Status		
Class I	5060	11.58
Class II	14450	33.06
Class III	22765	52.08
Class IV	1436	3.29
Previous CABG	16691	38.18
Diagnostics		
Abnormal conduction		
Normal	18513	42.35
LBBB	13671	31.28
Other	11527	26.37
Sodium		
<135	4583	10.48
135 to 145	38656	88.44

>145	472	1.08
Hemoglobin (5 g/Dl)	2.55	0.40
BUN (10 mg/Dl)	2.56	1.35
ICD Procedure(s)		
ICD Type		
Single Chamber	7716	17.65
Dual Chamber	17527	40.10
Biventricular	18468	42.25

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

The measure reliability dataset linked the ICD registry and Medicare Part A claims data from 2010Q2-2011Q4. The combined two-year sample included 43, 711 to 1,279 hospitals with 21,856 admissions to 1,254 hospitals in one randomly selected sample and 21,855 admissions to 1,246 hospitals in the remaining sample for patients aged 65 years and older. After excluding hospitals with fewer than 25 cases in each sample, the first sample contained 297 hospitals and the second sample contained 298 hospitals. In addition to being used for reliability testing, the linked dataset was used for measure exclusions testing (Section 2b3).

These analyses used a cohort of patients undergoing ICD placement for whom NCDR ICD Registry data were linked with corresponding administrative claims data. However, we also conducted additional analyses to meet newer testing requirements, and these analyses were performed using comparable linked data from 2010-2011. Details are provided below.

<u>Reliability testing (Section 2a2) and exclusions testing (Section 2b3)</u>
The measure reliability dataset linked the ICD and Medicare Part A claims data for claims between 2010Q2 and 2011Q4. The sample included ICD placement in a cohort of 43,711 Medicare FFS patients aged 65 years and older performed in 1279 hospitals. We then randomly split the sample, leaving 21,856 admissions to 1,254 hospitals in one randomly selected sample and 21,855 admissions to 1,246 hospitals in the remaining sample for patients aged 65 years and older. After excluding hospitals with fewer than 25 cases in each sample, the first sample contained 297 hospitals and the second sample contained 298 hospitals. The linked dataset was also used for measure exclusions testing (Section 2b3).

<u>Validity testing (Section 2b2)</u>

A summary of validity testing undertaken has been provided in 3c.1 of the application form. A chart validation study has been completed to determine whether ICD-9-CM diagnosis and procedure codes reported on Medicare claims and used in the measure specifications accurately identified patients experiencing ICD complications within 30 and 90 days of ICD implantation as reported in the medical charts. The findings of the study reported an overall agreement between chart and claims (based on paired ratings) of 91.5%.

Measure development and risk-adjustment dataset (Section 2b4)

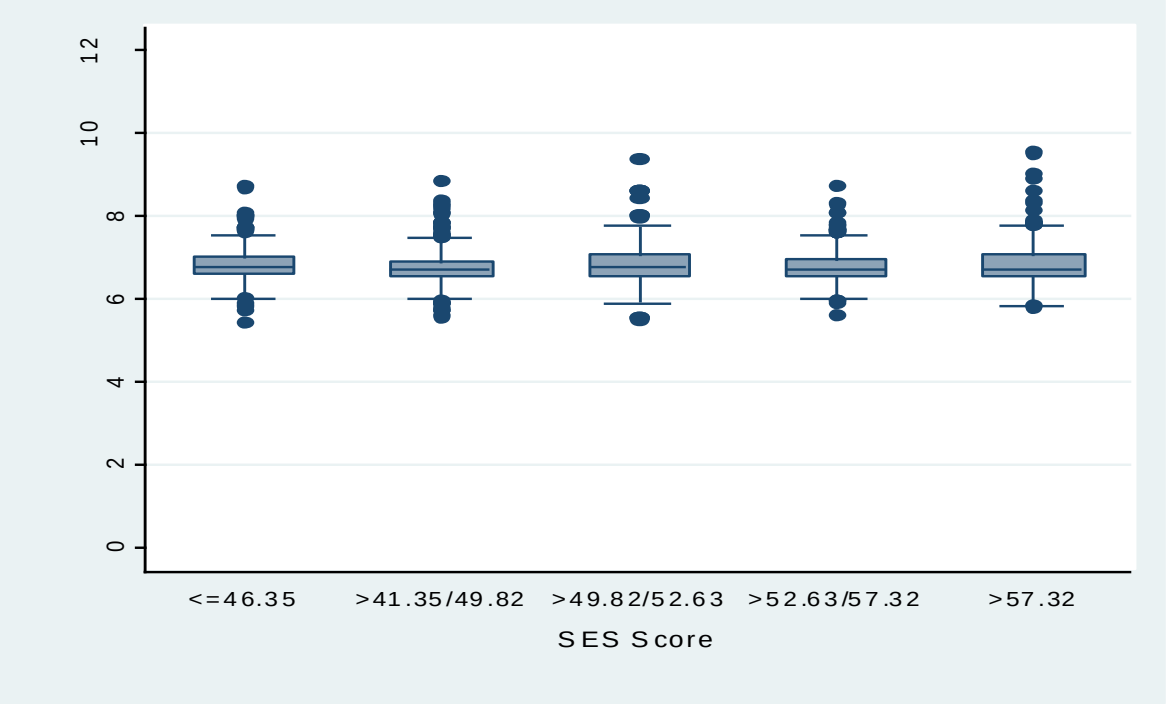
In measure development, we identified ICD procedures in the NCDR ICD Registry in which the patient was released from the hospital between April 2010 and December 2011. We merged ICD admissions in the NCDR ICD Registry data and ICD admissions in Medicare claims data to derive cohorts for development using probabilistic matching methodology. There were 21,855 cases discharged from the 1226 hospitals in the validation sample. This validation sample had a crude complication rate of 6.71%

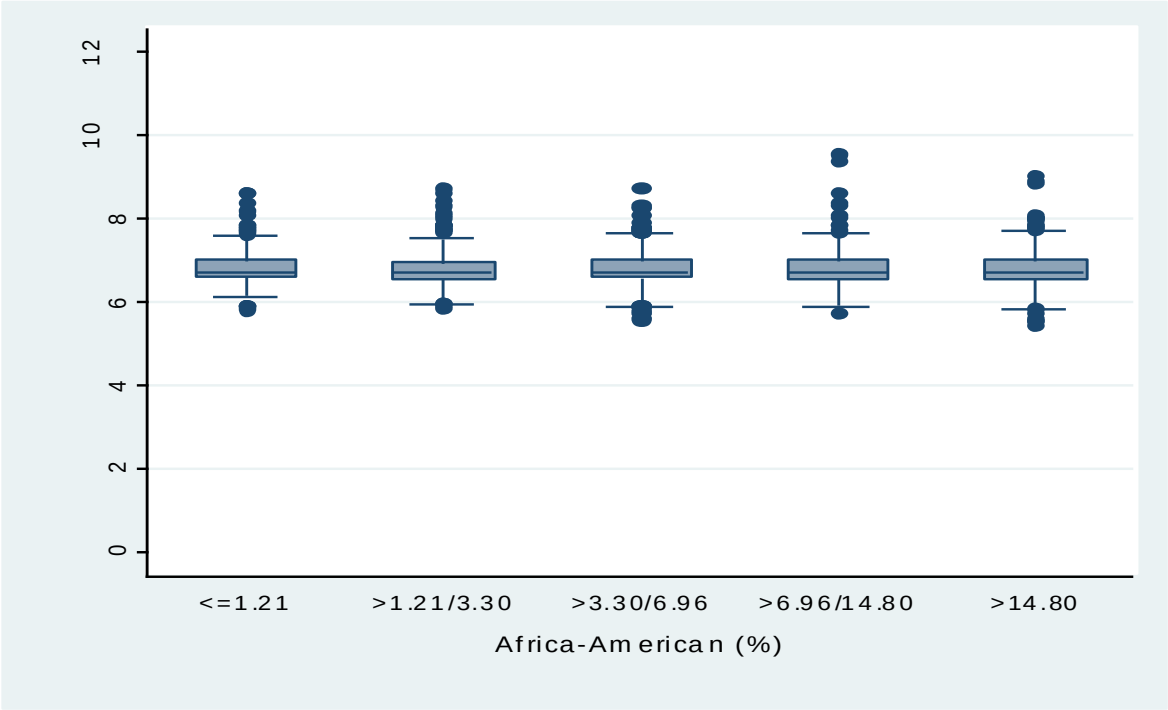
1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We examined health disparities associated with this measure by race/ethnicity and by socioeconomic status (SES). For both sets of analyses, the hospitals were grouped by quintiles of the rate of Africa Americans or the median SES score within a hospital. SES status (SES score) was determined from AHRQ SES Index data, and the percentage of African American patients was derived from CMS FFS patients for any condition during 2010 and 2011.

Dates of Data: 2010Q2 – 2011Q4
Number of Measured Entities (Hospitals): 1279
Number of Patients: 43711

The box and whisker plots below demonstrate the distribution of risk standardized complication rates (RSCRs) by SES score, and by the percentage of African-American patients.





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*)

Data Element Reliability

In constructing the measure we aim to utilize only those data elements from the claims that have both face validity and reliability. We avoid the use of fields that are thought to be coded inconsistently across hospitals or providers. Specifically, we use fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of CMS auditing and billing policies and seek to avoid variables which do not meet this standard. For example, "discharge disposition" is a variable in Medicare claims data that is not thought to be a reliable variable for identifying a transfer between two acute care facilities. Thus, we derive a variable using admission and discharge dates as a surrogate for "discharge disposition" to identify hospital admissions involving transfers. This allows us to identify these admissions using variables in the claims data which have greater reliability than the "discharge disposition" variable. In addition, CMS has in place several hospital auditing programs used to assess overall claims code accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in our measures, including diagnosis and procedure codes and other elements that are consequential to payment.

In addition, as an example of some of the methods that could be used to ensure data quality, we describe the NCDR's existing Data Quality Program (DQP). The two main component of the DQP are complementary and consist of the Data Quality Report (DQR) and the Data Audit Program (DAP). The DQR process assesses the completeness and validity of the electronic data submitted by participating hospitals. Hospitals must achieve >95% completeness of specific data elements identified as 'core fields' to be included in the registry's data warehouse for analysis. The 'core fields' include the variables included in 25 our risk adjustment models. The process is iterative, providing hospitals with the opportunity to correct errors and resubmit data for review and acceptance into the data warehouse. The DAP consists of annual on-site chart review and data abstraction. Among participating hospitals that pass the DQ random charts of 10% of submitted cases.

Finally, we assess the reliability of the data elements by comparing model variable frequencies and odds ratios in two years of data.

Measure Score Reliability

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

For test-retest reliability of the measure score, from the study cohort, we randomly sampled half of patients within each hospital, calculated the measure for each hospital in the first half, and then repeated the calculation using the second half. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement we calculated the intra-class correlation coefficient (ICC) (Shrout and Fleiss, 1979), and assessed the values according to conventional standards (Landis and Koch, 1977). Specifically, we used the two data samples and calculated the risk-standardized complication rate (RSCR) for each hospital for each sample. The agreement of the two RSCRs was quantified for hospitals in each sample using the intra-class correlation (ICC) as defined by Shrout and Fleiss (1979).

Using two independent samples provides an honest estimate of the measure's reliability, compared with using two random, but potentially overlapping samples, which would exaggerate the agreement. Moreover, because our final measure is derived using hierarchical logistic regression, and a known property of hierarchical logistic regression models is that small volume hospitals contribute less ´signal´. As such a split sample using a single measurement period likely introduces extra noise; potentially underestimating the actual test-retest reliability that would be achieved if the measures were reported using additional years of data. Furthermore, the measure is specified for the entire ICD population, but we tested it only in the subset of Medicare FFS patients for whom information about vital status was available.

References:

1) Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test–retest reliability of continuous measurements. Statistics in Medicine 2002;21:3431-3446.
2) Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-428.

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

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*)

Data element reliability results

Overall, risk factor frequencies changed little across years, and there were no notable differences in the odds ratios across years of data.

Split Sample Methodology:

Distribution of RSCR within random split samples

Description	**First (RAND= 1)**	**Second (RAND= 0)**
N	1254	1246
Mean	17.43	17.54
Std. Deviation	18.01	17.99
100% Max	0.0839	0.0836
75% Q3	0.0713	0.0676
50% Median	0.0697	0.0653
25% Q1	0.0684	0.0634

Results of the split sample testing are provided above. The 2 split samples were calculated during the same timeframe to avoid the potential for changes in hospital performance over time. After splitting the cohort into two random samples, we compared measure scores calculated at hospitals with at least 25 cases in both random samples. The distribution of hospital performance was similar in the two samples (figure below), and there was a fair correlation between hospital performances assessed in the two samples (r 0.1494).

Measure score reliability results

In the most recent years of data (2010Q1-2011Q4), there were 43,711 admissions in the combined two-year sample, with 21,856 admissions to 1,254 hospitals in the first randomly selected sample (mean RSCR 7.01%), and 21,855 admissions to 1,246 hospitals in the second randomly-selected sample (mean RSCR 6.58%). The agreement between the two RSCRs for each hospital was 0.1494, which according to the conventional interpretation is “slight” (Landis & Koch, 1977). The intra-class correlation coefficient is based on a split sample of 2 years of data, resulting in a volume of patients in each sample equivalent to only 1 year of data, whereas the measure is likely to be reported with a full two years of data.

Reference.

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

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

The stability over time of the risk factor frequencies and odds ratios indicate that the underlying data elements are reliable. Additionally, the ICC score demonstrates fair agreement across samples using a "strict" approach to assessment that would likely improve with greater sample size.

References:
Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. Mar 1977;33(1):159-174.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
Measure validity is demonstrated through prior validity testing done on our other measures, through use of established measure development guidelines, by systematic assessment of measure face validity by a technical expert panel (TEP) of national experts and stakeholder organizations, and through registry data validation.

Validity of Registry Data
Data element validity testing was done on the specified measure by comparing with variables in the ACC audit program. The NCDR ICD Registry has an established DQP that serves to assess and improve the quality of the data submitted to the registry. There are two complementary components to the Data Quality Program- the Data Quality Report (DQR) and the Data Audit Program (DAP). The DQR process assesses the completeness of the electronic data submitted by participating hospitals. Hospitals must achieve >95% completeness of specific data elements identified as "core fields" to be included in the registry's data warehouse for analysis. The "core fields" encompass the variables included in our risk adjustment models. The process is iterative, providing hospitals with the opportunity to correct errors and resubmit data for review and acceptance into the data warehouse. All data for this analysis passed the DQR completeness thresholds.

The DAP consists of annual on-site chart review and data abstraction. Among participating hospitals that pass the DQR for a minimum of two quarters, at least 5% are randomly selected to participate in the DAP. At individual sites, auditors review charts of 10% of submitted cases. The audits focus on variables that are used in the NCDR risk-adjusted in-hospital mortality model including demographics, comorbidities, cardiac status, coronary anatomy, and ICD status. However, the scope of the audit could be expanded to include additional fields. The DAP includes an appeals process for hospitals to dispute the audit findings.

We also examined the temporal variation of the standardized estimates and frequencies of the variables in the development and validation models.

To assess the predictive ability of the model, we grouped patients into deciles of predicted 30 or 90 day complication and compared predicted complication with observed complication for each decile in the derivation cohort (figure 2).

As noted in more detail in 3c.1 of the application form, a chart validation study has been completed to determine whether ICD-9-CM diagnosis and procedure codes reported on Medicare claims and used in the measure specifications accurately identified patients experiencing ICD complications within 30 and 90 days of ICD implantation as reported in the medical charts. The findings of the study reported an overall agreement between chart and claims (based on paired ratings) of 91.5%. Table 2 and figure 1 provide a depiction of the agreement and disagreement for patients with complication identified by charts and claims.

Figure 1. Number of Patients with Complication Identified by Charts and/or Claims

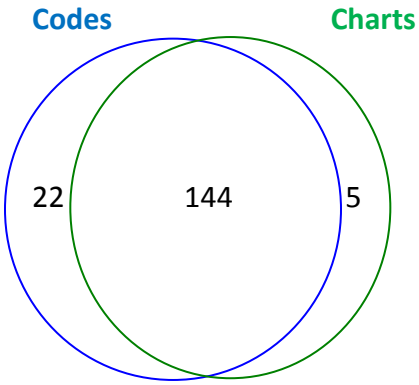


Table 2. Number of Cases of Disagreement and Agreement for Patients with Complication Identified by Charts and Claims

		CHARTS	
		YES	**NO**
CODES	**YES**	**144**	**22**
	NO	**5**	**145**

Validity as Assessed by External Groups

During original measure development and in alignment with the CMS Measures Management System (MMS), we released a public call for nominations and convened a TEP when originally developing the measure. The purpose of convening the TEP was to provide input and feedback during measure development from a group of recognized experts in relevant fields. The TEP represented physician, consumer, hospital, and purchaser perspectives, chosen to represent a diverse of perspectives and backgrounds.

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.
[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.
[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were initially identified using 2013 General Equivalence Mapping (GEM) software. We then enlisted the help of clinicians with expertise in relevant areas to select and evaluate which ICD-10 codes map to the ICD-9 codes currently in use for this measure. An ICD-9 to ICD-10 crosswalk is attached in field S.2b.(Data Dictionary or Code Table).

Lead clinical expert

Jeptha Curtis, M.D., Associate Professor of Medicine, Section of Cardiovascular Disease, Yale University

The full list of individuals involved are listed under the main application section Ad.1.

Table 3. 30 / 90 Day Complications Model Performance: Results Based on the GLM

Indices	**Derivation Sample**	**Validation Sample**
Year	2010-11	2010-11
RR Calibration (γ0, γ1)[1]	(0.00, 1.00)	(0.03, 1.02)
Discrimination- Adjusted R-Square[2]	0.07	0.06
Discrimination -Predictive Ability[3] (lowest decile %, highest	(4.05, 25.08)	(3.80, 23.80)
Discrimination – ROC	0.64	0.642
Residuals Lack of Fit (Pearson Residual Fall %)		
<-2	0.00	0.00
[-2, 2)	93.19	93.43
> 2	6.81	6.57
Model χ^2 [Number of Covariates][4]	328 [9]	355 [9]

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The performance of the derivation and validation samples is similar. The areas under the receiver operating characteristic (ROC) curve are 0.640 and 0.642, respectively, for the two samples. In addition, they are similar with respect to predictive ability. For the derivation sample, the predicted complication rate ranges from 3% in the lowest predicted decile to 14% in the highest predicted decile, a range of 11%. For the validation sample, the corresponding range is 3% to 14%, also a range of 11%.
Additionally, the frequencies and regression coefficients are fairly consistent over the two years of data. Also, there was excellent correlation between predicted and observed complications.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

_____ The audits conducted by the ACC support the overall validity of the data elements included in this measure. The data elements used for risk adjustment were consistently found for all patients and were accurately extracted from the medical record.

_____

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion. These exclusions are consistent with similar NQF-endorsed complication measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

2b3.2. 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*)

We examined overall frequencies and proportions of the admissions excluded for each exclusion criterion in the most recent data (2010Q1-2011Q4). The initial sample without exclusions included 112,341 admissions to 1,319 hospitals. After applying the exclusion criteria as outlined in Table 4, the 2010-2011 study sample included 43,711 patients admitted to 1,279 hospitals.

Table 4. Exclusions from the target cohort for the combined 2010-2011 study sample.

Exclusions	2010Q2 to 2011Q4			
	Patient Stay		Hospitals	
	#	%	#	%
Initial Sample	**112341**		**1319**	
Not Medicare patient on admission	14274	12.71	6	0.45
Remaining	**98067**		**1313**	
Not the first claim in the same claim bundle*	10	0.01	0	0.00
Remaining	**98057**		**1313**	
Not a full three months follow-up	1910	1.95	4	0.30
Remaining	**96147**		**1309**	
Previous ICD	44353	46.13	26	1.99
Remaining	**51794**		**1283**	
Previous pacemaker	8083	15.61	4	0.31
Study Sample	**43711**		**1279**	
Complication		0.00		
Mortality		0.00		
Mortality or Complications		0.00		

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The majority of exclusions are necessary to 1) link registry and administrative data (e.g. excluding patient not enrolled in Medicare FFS) and 2) identify patients eligible for complication (e.g. excluding patients who died before discharge). As such, these exclusions are not discretionary and do not require further testing.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
N/A. This measure is risk adjusted.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

We developed a parsimonious model that included key variables previously shown to be associated with complications following ICD implantation. Importantly, the variables included in the risk model were fully harmonized with the NCDR's existing risk model used to provide hospitals with risk adjusted in-hospital adverse events. In the development of that model, a team of clinicians had reviewed all variables in the NCDR ICD Registry database (a copy of the data collection form and the complete list of variables collected and submitted by hospitals can be found at www.ncdr.com and also in this application).

Based on clinical review informed by the literature, a total of 15 variables were determined to be appropriate for consideration as candidate variables. We used logistic regression with stepwise selection (entry p<0.05; retention with p<0.01) for variable selection. We also assessed the direction and magnitude of the regression coefficients. This resulted in a final risk-adjusted complication model that included 9 variables (table 5). To harmonize the models, we elected to apply this approach to risk adjustment to the 30/90 day complications risk model. Several variables were not clinically significantly associated with risk of complications at 30/90 days, but we elected to retain them in the model for consistency. We compared hospitals' RSCR calculated using this model with the output from a risk model that been developed specifically for 30/90 day complications and found them to be almost identical (correlation coefficient 0.996).

For categorical variables with missing values, the value from the reference group was added. The percentage of missing values for all categorical variables was very small (<1%) and they were imputed to a specific categories based on our previous experience. There were three continuous variables with missing values: **Hemoglobin** (HGB, 1.9%), **BUN** (1.3%), and **Sodium** (1.1%); and these missing values were imputed as the median of the non-missing values of the corresponding variable.

2b4.4. What were the statistical results of the analyses used to select risk factors?

Table 5. ICD Complication Model Variables

Description (n) Variable	**NCDR ICD Item Number (V2.1)**
1. Sex	2060
Male	
Female	
2. Reason for Admission	3010

Description (n) Variable	**NCDR ICD Item Number (V2.1)**
Admitted for procedure	
Cardiac Heart Failure	
Other	
3. NYHA Class	4020
I/II	
III	
IV	
4. Prior CABG	4190
5. Abnormal Conduction	5090
No	
Yes- left bundle	
Yes- other	
6. ICD Type	6130
Single Chamber	
Dual Chamber	
CRT-D	
7. Sodium	5125
<135	
135-145	
>145	
8. Hemoglobin (5g/Dl)	5120
9. BUN (10mg/Dl)	5115

Table 6. ICD Complication Model

9 variables	**Derivation cohort**		**Validation cohort**	
Effect	**Odds Ratio (95% CI)**	**P value**	**Odds Ratio (95% CI)**	**P value**
Sex				
Male	reference		reference	
Female	1.203 (1.068 - 1.354)	0.0022	1.218 (1.080 - 1.373)	0.0013
Reason for admission				
Admitted for procedure	reference		reference	
Cardiac heart failure	1.826 (1.560 - 2.137)	<.0001	1.559 (1.326 - 1.832)	<.0001
Other	1.967 (1.737 - 2.226)	.	1.932 (1.705 - 2.189)	.
NYHA class				
I/II	reference		reference	
III	1.183 (1.043 - 1.341)	0.0049	1.187 (1.045 - 1.349)	<.0001
IV	1.448 (1.112 - 1.885)	.	2.059 (1.618 - 2.620)	.
Prior CABG	0.839 (0.748 - 0.940)	0.0025	0.834 (0.743 - 0.936)	0.002
Abnormal conduction				
No	reference		reference	
Yes-left bundle	0.976 (0.837 - 1.137)	0.5575	0.962 (0.823 - 1.124)	0.462
Yes-other	1.054 (0.918 - 1.209)	.	1.055 (0.919 - 1.211)	.
ICD type				
Single chamber	reference		reference	
Dual chamber	0.877 (0.731 - 1.051)	0.3377	0.945 (0.785 - 1.139)	0.129

CRT-D	0.965 (0.833 - 1.117)	.	1.101 (0.950 - 1.277)	.
Sodium				
<135	1.300 (1.116 - 1.514)	0.0034	1.211 (1.037 - 1.415)	0.0301
135-145	reference		reference	
>145	1.011 (0.612 - 1.671)	.	0.758 (0.430 - 1.336)	.
Hemoglobin (5 g/Dl)	0.752 (0.646 - 0.876)	0.0003	0.719 (0.618 - 0.837)	<.0001
BUN (10 mg/Dl)	1.104 (1.065 - 1.144)	<.0001	1.130 (1.090 - 1.171)	<.0001

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Approach to assessing model performance

During measure development, we computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the development and validation cohort:

Discrimination Statistics:

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

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

Calibration Statistics:

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

We compared the model performance in the development sample with its performance in another sample of half of the patients randomly selected from the whole 2010Q2-2011Q4 study cohort. There were 21, 856 cases discharged from the 1222 hospitals in the 2010-2011 validation dataset. This validation sample had a crude complication rate of 6.86%. We also computed statistics (1) and (2) for the current measure cohort, which includes discharges from 2010Q2-2011Q4.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

For the derivation cohort the results are summarized below:
C-statistic=0.640
Predictive ability (lowest decile %, highest decile %): 4.05%, 25.08%

For the validation cohort the results are summarized below:
C statistic=0.642
Predictive ability (lowest decile %, highest decile %): 3.80%, 23.80%

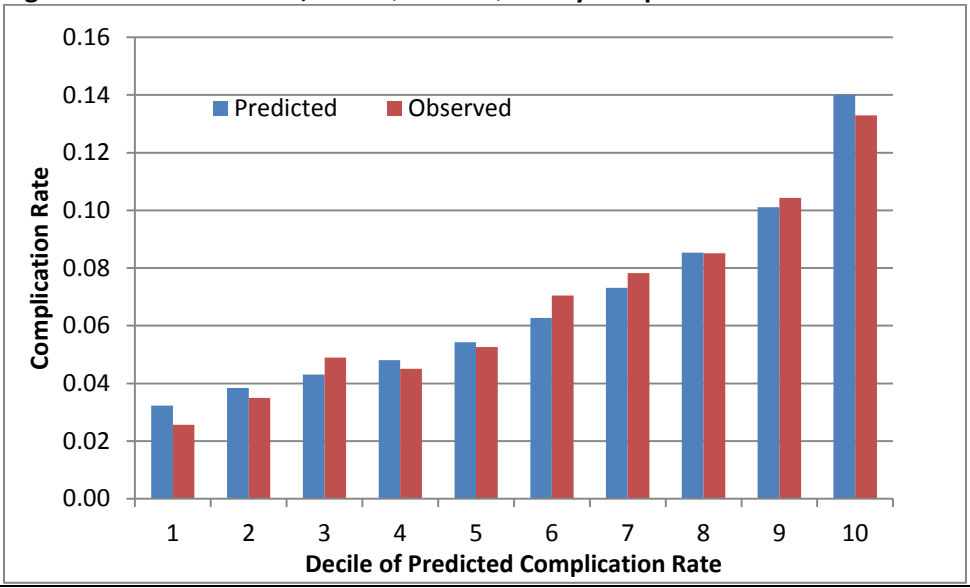
2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

For the validation cohort the results are summarized below:
Calibration: (0.03, 1.02)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

The risk decile plot is a graphical depiction of the deciles calculated to measure predictive ability. Below, we present the risk decile plot showing the distributions for the current measure cohort.

Figure 2. Risk Decile Plot, 2010Q2-2011Q4 study sample.



2b4.9. Results of Risk Stratification Analysis:

n/a

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

Discrimination Statistics
The C-statistics of 0.640 and 0.642 indicate fair model discrimination in the derivation and validation cohorts. Complications, as opposed to other outcomes such as mortality consistently have a lower c-statistic, even in medical record models. This is likely because complications are less determined by patient comorbidities and more by health system factors. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk patients from low-risk patients.

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

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

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

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

N/A

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. 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)*

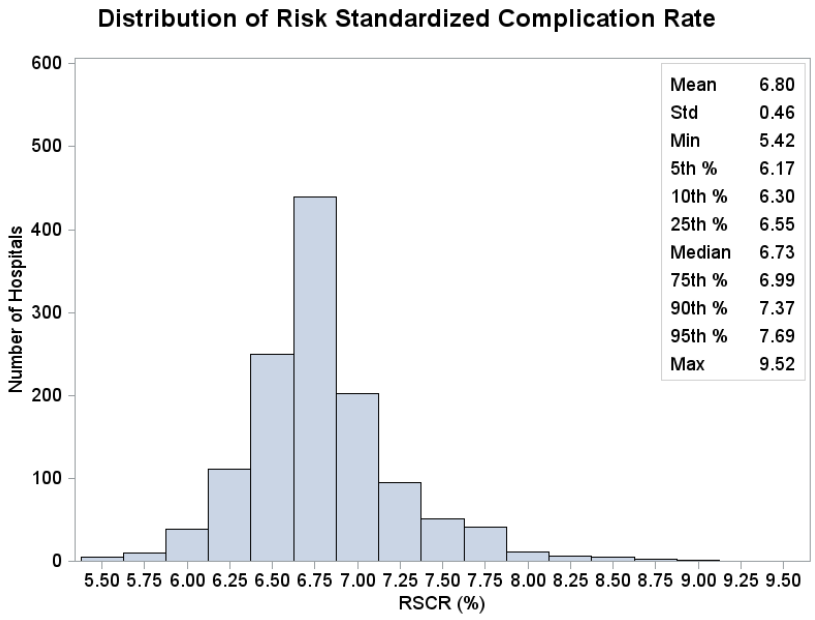
For the currently publicly reported measures of hospital outcomes, including the ICD 30/90 day complication measure, CMS estimates an interval estimate for each risk-standardized rate to characterize the amount of uncertainty associated with the rate. It then compares the interval estimate to the national crude rate for the outcome and categorizes hospitals as “better than,” “worse than,” or “no different than” the U.S. national rate (NCDR registry rate for ICD). We assessed variation in RSCRs among hospitals by examining the distribution of the hospital RSCRs and plotting the histogram of the hospital RSCRs.

2b5.2. 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*)

Recent analyses of Medicare FFS data show variation in RSCRs among hospitals. Using the most recent data sample (2010Q2-2011Q4) and updating the measure by applying the complication algorithm, the mean hospital RSCR was 6.8%, with a range of 5.42% to 9.52%. The interquartile range was 6.55% to 6.99%.

Figure 3. Distribution of risk-standardized complication rates (RSCRs); 2010Q2-2011Q4 sample.

Figure 3.



2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

The variation in rates suggests there are clinically meaningful differences across hospitals for 30/90 day risk-standardized complications after ICD insertion.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
***If only one set of specifications, this section can be skipped*.**

Note*: This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

N/A . This measure has only one set of specifications.

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A . This measure has only one set of specifications.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A . This measure has only one set of specifications.

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A . This measure has only one set of specifications.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

We examined rates of missing data for all candidate variables and examined histograms of the frequency of missing data by hospital.

2b7.2. 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*)

Overall, the percentage of missing values for all categorical variables was very small (<1%) and they were imputed to a specific categories based on our previous experience. There were three continuous variables with missing values: **Hemoglobin** (HGB, 1.9%), **BUN** (1.3%), and **Sodium** (1.1%); And these missing values were imputed as the median of the non-missing values of the corresponding variable.

2b7.3. 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*)

As noted above, model performance was comparable when we included or excluded cases with missing data.

Additional analyses and input from the Developers of measure 0694:

Indicators of socioeconomic status are increasingly considered potentially important variables for inclusion in efforts to risk adjust outcomes measures. This is something that has been carefully considered by the stewards of this measure, and upon clinical review from the expert panel, found to be less associated with complications.

The routine inclusion of SDS variables into risk models has the potential to explain away meaningful and actionable differences in hospital performance. Analyses have shown that many hospitals caring for a higher proportion of disadvantaged patients still perform well on the measure. Furthermore, literature shows that the inclusion of SDS does not consistently meaningfully improve the discriminatory capacity of risk adjustment models, shown in urban areas with large disparities in wealth and income (Blum et al., 2014) and even more nationally representative samples (Eapen et al., 2015).

The variables considered for inclusion in the risk model were carefully considered for both clinical significance and strength of association with the primary outcome. Based on clinical review informed by the literature, a total of 15 variables were determined to be appropriate for consideration as candidate variables for the risk model. The final risk-adjusted complication model that included 9 variables was decided on after further logistic regression with stepwise selection for variable selection, as well as assessing the direction and magnitude of the regression coefficients. Importantly, as mentioned in section 2b4.2 of the testing form, variables incorporated in the risk model were fully harmonized with the NCDR's existing risk model used to provide hospitals with risk adjusted in-hospital adverse events.

The Agency for Healthcare Research and Quality (AHRQ)-validated SES index score provides an analytic approach to accounting for SDS status. It incorporates 7 independently weighted socioeconomic variables under 5 domains; occupation, income, wealth, education and housing. The ICD 30/90 day complications measure utilizes this score to show the distribution of risk standardized complication rates (RSCR) by SES status. Section 1.8 of the testing form details the results of this analysis, showing little appreciable variation in distribution of risk standardized complication rates. The conceptual reason and justification for not including SDS status as a variable within the model is demonstrated by these results. In these results, adjustment for socioeconomic status does not have a statistically meaningful impact on the complication measure results.

Sources:

Blum, A. B., Egorova, N. N., Sosunov, E. A., Gelijns, A. C., DuPree, E., Moskowitz, A. J., Keyhani, S. (2014). Impact of socioeconomic status measures on hospital profiling in New York City. *Circ Cardiovasc Qual Outcomes, 7*(3), 391-397.

Eapen, Z. J., McCoy, L. a, Fonarow, G. C., Yancy, C. W., Miranda, M. L., Peterson, E. D., Hernandez, A. F. (2015). Utility of Socioeconomic Status in Predicting 30-Day Outcomes After Heart Failure Hospitalization. *Circulation Heart Failure*, *114*, 473–480.

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

On August 18th, 2015, the developer was informed by NQF staff that measure 0694 should be considered a composite measure. Developers updated the forms where possible. However, no empirical analysis could be performed in the time provided.

We believe the content validity of this measure has been achieved by virtue of the noted expertise of those individuals who developed this measure. The individual components of the composite have already shown to

impact clinical outcomes and specific information on how ACC determined which complications to include is outlined in 1.d.2.

However the empirical analysis demonstrating the individual component measures fit the overall quality construct is currently being researched. The testing will focus on construct validation which will test the hypothesis on the theory of the construct that following these processes for patients with ICD implantations lead to better outcomes. This research is expected to ultimately be published in the medical literature.

Measuring and reporting complications fit the overall construct. Also give the data burden, we believed we achieved the object of parsimony by including as few elements as possible without impacting the psychometric properties of measure 0694.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

All complications, as clinically defined by the subject matter experts were used with equal weighting and this straightforward approach is based on the current electrophysiology research.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc*.; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

We can provide data once testing is completed at the next maintenance review.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

We can provide data once testing is completed at the next maintenance review.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

We can provide data once testing is completed at the next maintenance review.

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

We can provide data once testing is completed at the next maintenance review.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

While no empirical analysis was performed, the developer carefully reviewed the weighing rules. In essence the goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with

information about hospital-level, risk-standardized procedural complication rates following hospitalization for an ICD implantation. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as operator and hospital procedural expertise, communication among providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This measure was developed to identify institutions whose performance is better or worse than would be expected based on their patient case mix, and therefore promote hospital quality improvement and better inform consumers about care quality. In addition to mortality, all identified complications in the specifications were considered serious enough. If a patient had one or more complications, they get counted as a "complication".

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

All complications are weighted in the same manner.

2d3. Empirical analysis demonstrating that the approach for handling missing data minimizes bias (*i.e., achieves scores that are an accurate reflection of quality*)**.**

Note: *Applies to the overall composite measure; the focus is on missing data rather than exclusions, which are considered in 2b3.*

2d3.1. What is the overall frequency of missing data and the distribution of missing data across providers?

Please see section 2b7.2.

2d3.2. Describe the method used to compare approaches for handling missing data (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

We can provide data once testing is completed at the next maintenance review.

2d3.3. What were the statistical results obtained from the analysis of missing data? (e.g., *results of sensitivity analysis of effect of various rules for missing data; if no empirical analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

N/A

2d3.4. What is your interpretation of the results in terms of demonstrating that the approach used for missing data minimizes bias? (i.*e., what do the results mean in terms of supporting the selected approach for missing data; if no empirical analysis, provide rationale for the selected approach for missing data*)

We can provide data once testing is completed at the next maintenance review.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

The National Quality Forum (NQF) first endorsed this measure in September 2010. This measure provides information on the hospital risk-standardized rates of complications following the implantation of an ICD in Medicare fee-for-service (FFS) patients at least 65 years of age. In developing the measure, we used clinical data from the National Cardiovascular Data Registry (NCDR) ICD Registry for risk adjustment, and linked that data to CMS administrative claims data. The claims data were used to identify ICD-related complications via ICD-9-CM diagnosis and procedure codes.

Considering that the administrative database may be subject to coding errors and variation in coding practices within and across care settings, the ICD measure development team at YNHHSC/CORE chose to conduct a chart validation study. The goal was to determine whether ICD-9-CM diagnosis and procedure codes reported on Medicare claims and used in the measure specifications accurately identify patients experiencing ICD complications within 30 or 90 days of ICD implantation as reported in the medical charts. This approach required obtaining medical records of patients who had an ICD implanted from participating hospitals, abstracting data related to ICD complications (including number of complications, timing, severity and treatment), conducting a head-to-head comparison of data between Medicare claims and medical records to assess the degree of agreement, and finally, where appropriate, adjusting the list of codes and/or the cohort definition in the ICD measure specifications to improve the agreement.

We calculated the sample size requirement based on the desired degree of agreement between medical records and claims data, which can be categorized as fair, moderate, substantial, or almost perfect depending on the magnitude of the kappa coefficient2. Our initial calculation was based on achieving a “substantial” degree of agreement, and accounting for a within-hospital correlation coefficient of 0.03. This would have required approximately 860 medical records. To compensate for missing charts, we added 10-20 percent to the required sample size. This increased the required number to nearly 1000 charts, which would have resulted in

doubling our budgetary allotment. We therefore decided to keep the sample size at 500 medical records from 9 candidate hospitals. This sample size would allow a qualitative assessment of the ICD-9-CM codes used in the claims model while meeting NQF review committee standards, considering budgetary and time constraints.

Thus, our approach was to recruit 9 hospitals and request copies of medical charts for approximately 60 Medicare FFS patients who had an ICD implanted between 2005 and 2007, 30 with and 30 without complications at each hospital. This number also accounted for 10 to 20 percent of medical records that may be missing. Given the low ICD complication rate, we selected sites that had a minimum of 25 cases with complications over the three-year period.

Although we planned to review approximately 540 charts from 9 hospitals, the final sample size was 411. One of the 9 hospitals withdrew from the project and did not provide the charts, though it previously signed both, a data use agreement (DUA) and a business associate agreement (BAA). Because they withdrew 9 months into the study, there was insufficient time remaining in the contract to recruit a replacement hospital. The remaining 8 hospitals provided 411 charts out of the 480 requested, with 69 (14.4%) missing charts.

Summary of Study and Findings:

- 9 Hospitals agreed to participate in the study; 8 completed the study and one withdrew
- Of the 480 charts requested, 411 were obtained from 8 hospitals (14.4% of missing charts)
- Charts were abstracted by professional chart abstractors at an independent company, Information Collection Enterprises, LLC (ICE)
- We excluded 95 cases because they had a previous ICD (consistent with current measure specifications). Thus, the study was based on 316 patients
- We identified 149 patients with complications (1 or more complication) in the chart, while administrative codes identified 166 patients with complications; 144 patients with complications were identified by both charts and codes; 22 by codes only; and 5 by charts only.

- These findings resulted in an overall agreement between chart and claims (based on all paired ratings) of 91.5% [Yes/Yes(144) + No/No(145) / total (316)=0.915] with a kappa coefficient of 0.83 (0.7865 – 0.8907) which is in the “almost perfect” range. A depiction of the overall agreement can be found in section 2b2.2 of the testing form (figure 1. and table 2).

We examined all cases of disagreement between the charts and the claims for each complication. Complications reported in the charts but not identified in the claims were due to either missing codes from our measure specifications or a failure to report the complication in the claims (e.g. evacuation of a hematoma). Examination of cases where the complication was reported in the claims but not in the charts, revealed that the complication was not related to the ICD but to another procedure, device, or medical condition. Based on these results, we made the following changes to the cohort and definitions of complications:

- Added the following administrative claim codes to the measure specifications to capture more mechanical complications with revision: 996.72; 39.94; 37.77

- Excluded patients with a previous pacemaker from the cohort, considering the lack of availability of present-on-admission codes at this time

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

The ACCF’s program the National Cardiovascular Data Registry (NCDR) provides evidence based solutions for cardiologists and other medical professionals committed to excellence in cardiovascular care. NCDR hospital participants receive confidential benchmark reports that include access to measure macro specifications and micro specifications, the eligible patient population, exclusions, and model variables (when applicable). In addition to hospital sites, NCDR Analytic and Reporting Services provides consenting hospitals’ aggregated data reports to interested federal and state regulatory agencies, multi-system provider groups, third-party payers, and other organizations that have an identified quality improvement initiative that supports NCDR-participating facilities. Lastly, the ACCF also allows for licensing of the measure specifications outside of the Registry.

It should be noted that the centers already have to participate in this specific registry for reimbursement purposes so that currently almost all hospitals that implant ICDs in Medicare populations already participate. Hence there is no additional cost.

Measures that are aggregated by ACCF and submitted to NQF are intended for public reporting and therefore there is no charge for a standard export package. However, on a case by case basis, requests for modifications to the standard export package will be available for a separate charge.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. **Current and Planned Use**

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

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above, provide:**

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

N/A, not being publicly reported.

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

N/A, not being publicly reported.

4a.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.*)

ACC is committed to implementing this measure. ACC is an authorized organization to receive CMS data through the ResDAC application process. Unfortunately, it has been determined by ResDAC that this authorization does not permit use of CMS for performance measure reporting purposes, either to hospitals or for public display. ACC is currently in process of applying to be a Qualified Entity. It is unclear if this pathway will permit measure implementation. ACC also is commenting on and tracking proposed language in 21st Century Cures legislation, which does appear to create a pathway for use of CMS data for this type of reporting purpose.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Due to the changes to this measure since its endorsement, a direct comparison of results of the measure over time would not provide a true reflection of the extent of improvement. The development of the parsimonious model is described in more detail in section 2b4.3 of the testing form as well as the process of selection of clinically significant variables in the updated model.

4b.2. **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.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
As noted earlier, publicly reporting hospital risk-standardized ICD complication rates requires that the data submitted by hospitals be complete, consistent, and accurate. A protocol that assures accurate data for public reporting should be established prior to implementation. Steps to ensure data quality could include monitoring data for variances in case mix, chart audits, and possibly adjudicating cases that are vulnerable to systematic misclassification.

As an example of some of the methods that could be used to ensure data quality, we describe the NCDR's existing Data Quality Program (DQP). The two main components of the DQP are complementary and consist of the Data Quality Report (DQR) and the Data Audit Program (DAP). The DQR process assesses the completeness and validity of the electronic data submitted by participating hospitals. Hospitals must achieve >95% completeness of specific data elements identified as 'core fields' to be included in the registry's data warehouse for analysis. The 'core fields' capture many of the variables included in our risk adjustment models. The process is iterative, providing hospitals with the opportunity to correct errors and resubmit data for review and acceptance into the data warehouse. The DAP consists of annual on-site chart review and data abstraction. Among participating hospitals that pass the DQR for a minimum of two quarters, at least 5% are randomly selected to participate in the DAP. At individual sites, on-site auditors review charts of 10% of submitted cases. The NCDR audit focuses on variables used to determine whether patients meet accepted criteria for ICD implantation. However, the scope of the audit could be expanded to include additional fields. The DAP includes an appeals process that allows hospitals to reconcile audit findings.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

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

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
HRS is expected to submit a complications measure that is attributable at the physician level. ACC and HRS staff have been in close contact and the specifications should mirror in both consensus standards applications.

5a. Harmonization

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 completely harmonized?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Yes. ACC and HRS have met and ensured the specifications are aligned as closely as possible. The inclusion and exclusion criteria are identical.

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.)
HRS is expected to submit a complications measure that is attributable at the physician level. ACC and HRS staff have been in close contact and the specifications should mirror in both consensus standards applications.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** ICD_Complications_measure_Final-635703207983177210.xlsx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American College of Cardiology
Co.2 **Point of Contact:** Jensen, Chiu, comment@acc.org, 202-375-6000-
Co.3 **Measure Developer if different from Measure Steward:** American College of Cardiology
Co.4 **Point of Contact:** Jensen, Chiu, comment@acc.org, 202-375-6000-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
For this particular topic those individuals who were involved in identifying the key attributes and variables for this outcome measure were leaders and experts in the field of electrophysiology and data management. Serial phone calls were held to both define the eligible population and given process. These clinical leaders are noted below.

Working Group:
Jim Beachy, RCIS
Barbara Christensen, R.N., M.H.A.
Susan Fitzgerald, R.N., M.B.A.
Stephen Hammill, M.D.
Paul Heidenreich, M.D.
Kathleen Hewitt, R.N., M.S.N., C.P.H.Q.
Alan Kadish, M.D.
Christie Lang
Isabelle LeBlanc

Frederick Masoudi, M.D.
Kristi Mitchell, M.P.H.
Kathy Pontzer
John Rumsfeld, M.D., Ph.D
Lara Slattery, M.H.S.
John Spertus, M.D., M.P.H.
Al Woodward, Ph.D., M.B.A.

Technical Expert Panel:
Francis Ferdinand, M.D.
Ziad Issa, M.D.
Neil Jensen, MHA, MBA
Alan Kadish, M.D.
Bradley Knight, M.D.
Bruce Koplan, M.D., M.P.H.
Frederick Masoudi, M.D.
John Onufer, M.D.
Russell Robbins, M.D., MBA
John Rumsfeld, M.D., Ph.D
Andrea Russo, M.D.
Stuart Winston, D.O.

NCDR Clinical Subworkgroup ensured the measure demonstrated an opportunity for improvement, had strong clinical evidence, and was a reliable and valid measure. These members included the below individuals: Drs. Jeptha Curtis (Chair), Frederick Masoudi, John Rumsfeld, Matt Reynolds, and Mark Kremers.

NCDR Scientific Quality and Oversight Committee—a committee that served as the primary resource for crosscutting scientific and quality of care methodological issues. These members included Drs. Frederick Masoudi (Chair) , David Malenka, Thomas Tsai, Matthew Reynolds, David Shahian, John Windle, Fred Resnic, John Moore, Deepak Bhatt, James Tcheng, Jeptha Curtis, Paul Chan, Matthew Roe, and John Rumsfeld.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2011

Ad.3 **Month and Year of most recent revision:** 02, 2015

Ad.4 **What is your frequency for review/update of this measure?** With dataset revisions and based on new evidence.

Ad.5 **When is the next scheduled review/update for this measure?** 02, 2016

Ad.7 **Disclaimers:** ACC realizes the various NCDR endorsed measures are not readily available on their own main webpage. However, ACCF plans to update their main webpage (acc.org) to include the macrospecifications of the NQF endorsed measures. ACC hopes to work collaboratively with NQF to create a consistent and standard format would be helpful for various end users. In the interim, the supplemental materials include the details needed to understand this model. In addition, interested parties are always able to contact comment@acc.org to reach individuals at the ACC Quality Measurement Team.

Ad.8 **Additional Information/Comments:** ACC appreciates the opportunity to submit measures for this NQF endorsement maintenance project.

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 #: 0730
Measure Title: Acute Myocardial Infarction (AMI) Mortality Rate (IQI15)
Measure Steward: Agency for Healthcare Research and Quality
Brief Description of Measure: In-hospital deaths per 1,000 hospital discharges with acute myocardial infarction (AMI) as a principal diagnosis for patients ages 18 years and older.
Developer Rationale: Acute Myocardial Infarction (AMI) is an emergent condition for which significant morbidity and mortality may result from delayed, inappropriate or low quality treatment. Processes that have been identified to lower AMI mortality include: timely electrocardiography (ECG), early percutaneous coronary intervention (PCI) for ST-elevation MI, or fibrinolytic therapy if PCI is not available, aspirin administration at arrival, afterload reduction with renin-angiotensin system blockade among patients with impaired left ventricular function, and aspirin/statin prescription initiation during hospitalization with continuation at discharge. The 30-day risk-standardized AMI mortality rate is currently a CMS quality metric. In-hospital risk-adjusted mortality rates allow users without access to post-hospitalization death data, to assess and compare AMI outcomes.

Numerator Statement: Number of in-hospital deaths among cases meeting the inclusion and exclusion rules for the denominator.
Denominator Statement: Discharges, for patients ages 18 years and older, with a principal ICD-9-CM diagnosis code for AMI.
Denominator Exclusions: Exclude cases:

- transferred to another short-term hospital, for whom the outcome at hospital discharge was unknown
- admitted for treatment of pregnancy, childbirth, and puerperium
- with missing discharge disposition, gender, age, quarter, year, or principal diagnosis

Measure Type: Outcome
Data Source: Administrative claims
Level of Analysis: Facility
Is this an eMeasure? ☐ Yes ☒ No **If Yes, was it re-specified from a previously endorsed measure?** ☐ Yes ☐ No

Is this a MAINTENANCE measure submission? ☒ Yes ☐ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: Mar 28, 2011 **Most Recent Endorsement Date:** Mar 28, 2011

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (National Voluntary Consensus Standards for Patient Outcomes 2009):

This measure provides a rate of in-hospital AMI mortality using administrative data. It was compared to another endorsed in-hospital AMI mortality measure from The Joint Commission (161 AMI inpatient mortality). The Joint Com- mission is no longer reporting their in-hospital AMI mortality measure5 on their website in favor of CMS's NQF-endorsed 230 AMI 30-day mortality measure. This candidate AMI mortality measure from AHRQ differs from measure 161 in that the risk-adjustment model is based on all patient refined diagnosis related groups (APR DRGs), uses administrative coding rather than manual medical record abstraction, and does include transfers into the facility. Reliability of the coding was demonstrated to be 93 percent to 98 percent. The population measured is determined by the principal diagnosis, and the definition of AMI is harmonized with the endorsed 30-day AMI mortality measure from CMS. The Committee considered the differences in the measures and the benefits of having both inpatient and 30-day mortality measures. Unlike the 30-day mortality measure, which includes only patients aged >65 years, this candidate standard includes all patients experiencing AMI as a primary diagnosis. The inpatient measure is more feasible for some implementers since tracking out-of-hospital deaths can be difficult. Members of the Steering Committee also felt that knowing the proportion of in-hospital deaths was important in addition to the 30-day mortality data and that the two measures are complementary. Committee members asked the developers about the 30 percent of AMI patients who are excluded with a secondary AMI diagnosis and are not

captured in the measure currently. The developer clarified that most excluded patients experienced an AMI postoperatively, and the Committee suggested that future measures should address this population.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if health outcomes measures agree the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

- This is a risk-adjusted (by gender, age & clinical co-morbidities) outcomes measure that assesses the occurrence of in-hospital deaths per 1,000 hospital discharges with acute myocardial infarction (AMI) as a principal diagnosis for patients ages 18 years and older. The developer states there are numerous health care processes that lower AMI mortality.
- This measure is a component for the Inpatient Quality Indicators #91 (IQI #91) Mortality for Selected Conditions measure, which also includes individual component measures for Heart Failure, Acute Stroke, Gastrointestinal Hemorrhage, Hip Fracture and Pneumonia Mortality Rates.
- In addition to demonstrating links between processes and outcomes, the developer provides numerous clinical practice guidelines for the evaluation, management and treatment of AMI, as well as a formalized environmental scan of AMI literature related to health system characteristics and processes, relationship to 30-day mortality, geographic and temporal variation, risk adjustment, and scientific acceptability.

Questions for the Committee (as appropriate):

- *Are there structures or processes of care that can affect this outcome?*

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

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

- The developer provided a robust reference population demonstrating an improvement opportunity of overall AMI inpatient mortalities per 1000 discharges (disposition = 20) for over 2800 hospitals with statistically significant declining annual mortalities of 68.94 to 56.37 from 2008 to 2012 using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID). For IQI 15, 24,890 patients died among 441,557 AMI patients at 2,978 hospitals in 2012. Observed rates from 2008-2010 are not provided as the AMI diagnosis data element was not "Present on Admission" (POA) or collected consistently during that time. In hospital complications, for risk adjustment purposes, are also excluded as they would not be POA. Distributions, mean, standard deviation (SD) of the observed rates are included, and 5th, 25th, 50th (median), 75th, and 95th percentile are also provided.
- 2012 HCUP SID patient characteristic disparities data are provided by age, gender, zip code median income, NCHS patient residence location, and expected payment source, with disparities increasing with age (expected), female, highest and lowest zip code median income, large central metropolitan and micropolitan residences, and Medicare payer. Hospital characteristic disparities data includes Northeast, Midwest, South & West locations of care, noting disparities in the Northeast and West regions.

Questions for the Committee:

○ *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

○ *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - There is excellent evidence for this Outcome Measure

1a. ***Committee's Comments on Evidence to Support Measure Focus:***
   - Not Applicable

1b. ***Committee's Comments on Performance Gap:***
   - There is a performance GAP and disparity for this measure.

1c. ***Committee's Comments on Composite Performance Measure:***
   - Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- This measure is specified at the facility level of analysis exclusively using electronic administrative claims as the data source to calculate the measure performance rate, with better quality equaling lower scores. The measure utilizes "all payer" discharges (disposition = 20) for the calendar year for patients ages 18 years and older, with a principal ICD-9-CM diagnosis code for AMI for all hospitals, except for psychiatric facilities, alcohol and drug dependency facilities and military hospitals. The data elements are clearly defined, and ICD-9 & ICD-10 codes are provided. A peer-reviewed ICD-9 and ICD-10 conversion methodology is discussed and was validated through literature review and IQI15 annual use review & measure maintenance activities. The developer utilizes patient age, sex, disposition, principal diagnosis for AMI, procedural, quarter and year codes to calculate the measure.
- The developer standardizes 4 provider-level indicators groups including demographics, severity of illness, comorbidities and transfer-in status for clinical factors covariates or potential risk adjusters.
- This outcome measure is risk-adjusted using a statistical risk model by gender, age, comorbidities and transfer in status. The calculation algorithm provided by the developer is informative and understandable. Additional details for the risk model are provided in the supplemental spreadsheet attachment.
- The measure is risk adjusted, and the developer provides descriptions and covariates for the risk variables used in the risk-adjustment model are included in the specifications (*gender, age, Major Diagnostic Categories (MDC), All Payer Refined Diagnosis Related Group (APR-DRG) with Risk of Mortality (ROM) scores, patient point-of-origin and whether they were transferred from another facility).* Other than gender and age, no other SDS factors were included in the risk-adjustment approach.
- Predicted values for each case are computed using logistic regression and hospital random effect, with the defined measure covariates. Descriptions of the reference population, expected rates, risk adjusted rates and observed rates are provided, with the risk adjusted rate calculated with the following equation:

Risk Adjusted Rate = Reference Population X (Observed Rate / Expected Rate)

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- The developer completed *performance measure score* testing using a signal-to-noise analysis, which is appropriate for this measure, which assesses differences in performance between hospitals ("the signal") to stability within hospitals (random measurement error or "the noise"). Hospital size is calculated by number of discharges per year, which is weighted to reduce impact of small denominators. The signal-to-noise method results in a reliability statistic that ranges from 0 to 1 for each facility. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in between hospital performance. A value of 0.7 is often regarded as a minimum acceptable reliability value, though the developer provides guidance for lower acceptable reliability scores.
- Performance score testing results are provided by deciles of 266 and 267 hospitals from 2012 HCUP SID data, with an overall signal-to-noise ratio of 0.75 for 2664 hospitals with an average of 165.6 discharges per year, and range from 0.169 for hospitals with 4.2 average discharges per year to 0.899 for hospitals with 672.2 average discharges per. Hospitals with greater than 28 discharges per year were risk adjusted 0.48 to .090 reliability.
- Individual cases not meeting denominator definitions, or with missing data described in 2b.7 of the Measure Worksheet are removed from the calculation.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- Because this is an outcome measure, the rationale that is presented for subcriterion 1a does not necessarily have to address all of the variables used to calculate the measure.
- The measure was initially released in 2003, has been in use since that time, and undergoes annual maintenance and implementation updates as needed.
- A literature review of potential SDS factors indicates that race, ethnicity, and income are associated with in-hospital mortality after AMI. However, this literature also suggests that these relationships may be mediated by the quality of care provided (e.g., higher door-to-balloon time for Black patients, lower utilization of circulatory support devices among poorer patients). Accordingly, the developer has chosen not to include these factors in their risk-adjustment approach.

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Does the Committee agree that the conceptual relationship between SDS factors and AMI in-hospital mortality as outlined by the developer does not warrant inclusion of SDS factors in the risk adjustment model? Are these variables available and generally accessible for the measured patient population?*

2b2. Validity testing

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.

- Validity testing was conducted with critical data elements, and performance measure score using empirical validity testing.
- Critical data element testing included a systematic review of alternative peer-reviewed definitions for AMI with ICD-9 coding, applied to a Canadian inpatient records from 2003, using ICD-9 & ICD-10 definitions resulting in the highest positive predictive value (PPV) of 84.0% and sensitivity of 81.1%. Seven studies validated the findings with patient data from 1984 through 2009.
- Empirical validity testing was performed between hospital-level Spearman rank correlation between IQI 15 risk-adjusted rates and adherence for 6 process measures; risk adjusted and risk standardized 30-day AMI mortality measures for all patients and for Medicare FFS beneficiaries; and 3 alternative cardiovascular care IQI measures with measure data for all measures from 2011 and/or 2012 generally demonstrating favorable correlations between the measures' performance and adjusted AMI mortality rates.
- As the measure has been in use for a more than 10 years, the developer states that along with data element validity testing and parallel finding and empirical validity testing from the literature cited, the measure's ability to differentiate quality of care between hospitals should be assumed.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Excluded patients include:
  - transferred to another short-term hospital, for whom the outcome at hospital discharge was unknown
  - admitted for treatment of pregnancy, childbirth, and puerperium.
  - with missing discharge disposition, gender, age, quarter, year, or principal diagnosis that is not "Present on Admission" (POA).

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- The risk model includes 23 risk factors under 4 covariate categories including gender, age, Major Diagnostic Categories (MDC), All Payer Refined Diagnosis Related Group (APR-DRG) with Risk of Mortality (ROM) scores, patient point-of-origin, and whether they were transferred from another facility. The developer describes the process for including risk factors based on AHRQ QI Empirical Methods Report, which outlines 4 steps in estimating risk modes (1. Construct candidate covariates, 2. Select model covariate, 3. Estimate the models, and 4. Evaluate the models)
- Measure discrimination assessing the risk model's ability to distinguish performance between hospitals is conducted with c-statistics with findings above 0.70 and below 0.80 have moderate discrimination, and above 0.80 as having high discrimination. Again, using 2012 HCUP SID inpatient data of approximately 44,000 discharges, a c-statistic of 0.8867, with deciles ranging between 0.85 – 1.04, and hospitals with fewer discharges resulting in lower predictive value. "Goodness of fit" tests were not performed due to large sample size, which is appropriate. The developer states the model has very strong predictive power.

The risk model includes clinical patient factors and does not include SDS factors beyond age and gender. The developer describes SDS factors, such as race/ethnicity, within the context of the measure focus; however, they note that these factors are confounded by the quality of care provided by the accountable entity.

Questions for the Committee:

For outcome measures:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*

2b5. Meaningful difference:

- Using smoothed indicator rates that assume each hospital has a Gamma distribution, 20th percentile benchmarks for the reference population, and 80th percentile thresholds, decile results are provided by hospital size (# of discharges) to demonstrate meaningful differences in performance. The results demonstrate the measure has strong discrimination for identify low performing and moderate to large hospitals above and below benchmarks.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- As there is only one data source used for measure calculation (administrative claims), comparability of data sources or methods is not applicable.

Question for the Committee:

-

2b7. Missing Data

- All state hospital data with 36 of 45 participating states (or ~82%) of national community hospital inpatient discharges are included for 2012 analysis.
- Patients with gender, age, quarter, year or principal diagnosis that are not "Present on Arrival" (POA) are excluded accounting for < 0.01%.

Questions for the Committee:

- *Do the missing records pose a threat to validity?*

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- Specifications are clearly defined

2a2.: ***Committee's Comments on Reliability-Testing:***

- No concerns with Reliability

2b1.: ***Committee's Comments on Validity-Specifications***:

- Not Applicable

2b2.: ***Committee's Comments on Validity-Testing:***

- No concerns with Validity

2b3-7.: ***Committee's Comments on Threats to Validity:***

- No Concerns

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than the person obtaining the original information (e.g., DRG, ICD-9 codes on claims).
- Users have over ten years of experience using the AHRQ QI software in SAS and Windows, which has been available at no cost since 2001. The IQI15 software is freely available from the AHRQ Quality Indicators website.
- This is not an eMeasure.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- Administrative Claims Data, no concerns

Criterion 4: Usability and Use

4. Usability and 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.

- The measure was first released in 2003 and is broadly used in public and private accountability and quality improvement programs, and is publically reported.
- The developer also states, when this measure is used in conjunction with 30-day risk-standardized AMI mortality rate, the in-hospital risk-adjusted mortality rates allow users without access to post-hospitalization death data, to assess and compare AMI outcomes.
- The developer found no noted unintended consequences, though theoretically expedited transfers to other levels of care could be found in dying patients. Though "transfer to another short-term hospital, for which the outcome at hospital discharge was unknown discharge" is a denominator exclusion, the developer states that in those instances, reporting the 30-day mortality rate measure would discourage such transfers.

Questions for the Committee*:*

- *For maintenance measures – is the measure used in at least one accountability application?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- Tested and in use for over 10 years, no concerns.

Criterion 5: Related and Competing Measures

0230 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older
2473 : Hospital 30-Day Risk-Standardized Acute Myocardial Infarction (AMI) Mortality eMeasure

The indicators referenced above include 30-day mortality 1) for patients age 18 years and older 2) specified as an e-measure and 3) for patients age 65 and older. Inpatient mortality and 30-day mortality are different concepts, although capturing the same ultimate outcome. Harmonization is not appropriate.

IQI 15 and the Centers for Medicare & Medicaid Services' NQF-endorsed measures concerning AMI mortality (0230 and 2473) use the same ICD-9-CM codes to identify AMI, but they differ in two important respects: (1) whereas the CMS measures concern only Medicare fee-for-service and VA beneficiaries 65 years or older, IQI 15 measures mortality among hospitalizations of patients 18 years or older at non-federal acute care hospitals for all payers; and (2) while the CMS measures evaluate 30-day mortality, IQI 15—because it is based only on UB-04 data elements—is limited to inpatient mortality. The latter difference is a potential disadvantage in that the time at risk is not uniform for all patients and 30-day mortality is typically greater than inpatient mortality, but the former difference is an advantage because IQI 15 encompasses a greater proportion of the entire population at risk. We therefore believe that #0730 complements #0230 by offering an alternative specification for users who are interested in patients of all ages and all payers, just as #2473 offers an alternative e-measure specification for those with electronic health data.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0730

Measure Title: Acute Myocardial Infarction (AMI) Mortality Rate (IQI15)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Inpatient Quality Indicators #91 (IQI #91)

Date of Submission: 6/30/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☒ Health outcome: mortality

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Acute Myocardial Infarction (AMI) is an emergent condition for which significant morbidity and mortality may result from delayed, inappropriate or low quality treatment. Processes that have been identified to lower AMI mortality include: timely electrocardiography (ECG), early percutaneous coronary intervention (PCI) for ST-

elevation MI, or fibrinolytic therapy if PCI is not available, aspirin administration at arrival, afterload reduction with renin-angiotensin system blockade among patients with impaired left ventricular function, and aspirin/statin prescription initiation during hospitalization with continuation at discharge. The 30-day risk-standardized AMI mortality rate is currently a CMS quality metric. In-hospital risk-adjusted mortality rates allow users without access to post-hospitalization death data, to assess and compare AMI outcomes.

Please see clinical practice guidelines cited below for specific evidence supporting this rationale statement.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Not applicable

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

Please note that this is an outcome measure, so a systematic review of the body of evidence that supports the performance measure is not required. However, information is provided in 1a.4.1 and 1a.8 below, to provide additional context and support for IQI 15.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

ACC/AHA 2007 guidelines for the management of patients with unstable angina/non–ST-elevation myocardial infarction—executive summary. 2007 Aug 14.

http://www.ncbi.nlm.nih.gov/pubmed/17692738

2011 ACCF/AHA Focused Update of the Guidelines for the Management of Patients With Unstable Angina/Non–ST-Elevation Myocardial Infarction (Updating the 2007 Guideline).

A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the American College of Emergency Physicians, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons.

http://content.onlinejacc.org/article.aspx?articleid=1146459

American College of Emergency Physicians. Clinical policy: critical issues in the evaluation and management of adult patients with non-ST-segment elevation acute coronary syndromes. Fesmire FM, Decker WW, Diercks DB, Ghaemmaghami CA, Nazarian D, Brady WJ, Hahn S, Jagoda AS, Ann Emerg Med 2006 Sep;48(3):270-301.

http://www.ncbi.nlm.nih.gov/pubmed/16934648

Myocardial infarction with ST-segment elevation. The acute management of myocardial infarction with ST-segment elevation. 2013 Jul. NGC:009974

National Clinical Guideline Centre (UK). London: Royal College of Physicians (UK); 2013 Jul.

National Institute for Health and Clinical Excellence: Guidance.

http://www.ncbi.nlm.nih.gov/pubmed/25340241

ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. 1996 (revised 2012 Oct). NGC:009544

http://www.ncbi.nlm.nih.gov/pubmed/22922416

2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction. A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. 1996 Nov 1 (revised 2013 Jan 29). NGC:009572

http://www.ncbi.nlm.nih.gov/pubmed/23256914

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not applicable

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not applicable

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

Not applicable

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not applicable

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

Not applicable

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 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.8.1 What process was used to identify the evidence?

Formal environmental scans of the literature, including routine PubMed searches are performed to continually update evidence. The current evidence review results presented below constitute the most recent update, conducted in January 2015. Search terms included relevant MeSH terms (Myocardial Infarction/mortality). We combined this clinical search string with MeSH terms (quality indicator, hospital mortality) to identify studies examining quality of inpatient care. The search was limited to English publications. For completeness we also tested more inclusive search strings. Below we have provided a summary of the most up-to-date evidence.

1a.8.2. Provide the citation and summary for each piece of evidence.

Association with health system characteristics and processes

Three studies examined the association between hospital teaching status and IQI15 rates. One found that major teaching hospitals in the same Medicare data set had 20% lower risk-adjusted 30-day mortality than nonteaching hospitals; half of this difference was attributable to greater use of beneficial therapies.[1] Two other studies found that in-hospital AMI mortality was significantly lower in major teaching hospitals than in minor and non-teaching hospitals.[1,2] A fourth study by Jena et al. used Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS) data (2002-2008) to examine the effect of resident physician turnover and subsequent influx of inexperienced new physicians in the month of July in teaching (n=98) and non-teaching hospitals (n=1353).[3] They found the adjusted in-hospital mortality rates of high-risk AMI patients in teaching hospitals to be lower in May than in July (18.8% vs. 22.7%, $p < 0.01$) but also found in-hospital mortality in the lower-risk AMI cohort to be similar in May and June for both teaching (2.1% vs. 1.9%, $p = 0.45$) and non-teaching (2.7% vs. 2.8%, $p = 0.21$) hospitals, respectively.

One study of Medicare data by Patel et al. found that there were no significant differences in 30-day mortality rates for AMI between more intensive vs. less-intensive teaching hospitals (odds ratio 1.05, 95% CI 0.97 to 1.14; p=0.20) following Accreditation Council for Graduate Medical Education (ACGME) resident work hour limit reforms in 2011.[4] Hospital teaching intensity in the study was represented by the change in odds of an event for every 1-unit change in the resident-to-bed ratio.

Multiple studies have examined the relationship between hospital volume and IQI15 rates. Carretta et al. used AHRQ IQI measures and 2008 patient-level discharge files (n=30,843 records of AMI inpatient mortality) from Florida to investigate the impact of hospital characteristics on IQI mortality rates and found that increased hospital volume (as measured by a log-transformed value) was associated with decreased in-hospital AMI mortality (OR 0.89; $p < 0.001$) and overall 30-day mortality (OR 0.91; $p < 0.0001$) while other hospital characteristics had little impact on AMI death outcomes.[5] Various studies also found that for primary percutaneous transluminal coronary angioplasty (PTCA), high volume hospitals had lower AMI mortality than lower volume hospitals.[6-12] For example, in the 1995 New York State Coronary Angioplasty Reporting System Registry, in-hospital mortality was reduced 57% among patients who underwent primary angioplasty by high-volume, as opposed to low-volume, physicians (adjusted relative risk 0.43; 95% CI 0.21 to 0.83). When patients with AMI were treated with primary angioplasty in high-volume hospitals rather than low-volume institutions, the relative risk reduction for in-hospital mortality was 44% (adjusted relative risk 0.56; 95% CI 0.29 to 1.1). When compared with patients treated at low-volume hospitals by low-volume physicians, patients treated at high-volume hospitals by high-volume physicians had a 49% reduction in the risk of in-hospital mortality (adjusted relative risk 0.51; 95% CI 0.26 to 0.99).[7]

Other studies have examined the impact of various factors such as hospital spending and accreditation. Analyzing NIS data (2003-2007) from over 1,200 hospitals, Romley et al found that patients treated at hospitals in the highest spending quintile had lower risk-adjusted inpatient mortality for AMI (OR 0.65; 95% CI 0.56 to 0.76) compared with hospitals in the lowest spending quintile.[15] Chen et al. examined the association between accreditation of hospitals by The Joint Commission, quality of care, and survival among Medicare patients hospitalized for AMI. They found that hospitals not accredited by The Joint Commission had, on average, lower quality (less likely to use aspirin, beta-blockers, and reperfusion therapy) and higher thirty-day mortality rates than accredited hospitals.[6]

Finally, several studies have explored associations between in-hospital risk-adjusted mortality rates and process adherence rates for AMI patients. Meehan et al. evaluated coding accuracy, severity of illness, and process-based quality of care in Connecticut hospitals. Three process measures were selected by an expert panel based on medical literature and local practice patterns: 1) administration of thrombolytic therapy, 2) discharged on aspirin if no contraindication, and 3) discharged on a beta blocker if no contraindication. Hospitals with the highest risk-adjusted mortality had significantly lower utilization of these therapies than other hospitals in the sample. Although the Medicare Prospective Payment System Quality of Care study did not focus on specific therapeutic interventions, it also demonstrated significantly higher risk-adjusted mortality (using risk factors derived by chart review) among hospitals with "poor" processes of care than among hospitals with "good" or "medium" processes of care (30.1% versus 22.0% and 23.9%, respectively). Chen showed that the hospitals designated by US News as "America's Best Hospitals" in cardiology, based on risk-adjusted in-hospital mortality (using APR-DRGs) and reputation among physicians, had lower risk-adjusted mortality using clinical predictors among Medicare patients (15.6% versus 18.3-18.6%), and used aspirin and beta blockers more often, than hospitals that were not so designated. Similar findings were reported using HealthGrades' approach to estimating risk-adjusted in-hospital mortality.[13] In another study, quality improvement interventions lowered the risk of in hospital death in patients with AMI about 40%. In the California Hospital Outcomes Project, hospitals with low risk-adjusted AMI mortality were more likely to give aspirin within 6 hours of arrival in the emergency department, more likely to perform cardiac catheterization and revascularization within 24 hours, and more likely to give heparin to prevent thromboembolic

complications. However, there were no differences between low and high-mortality hospitals in the use or timing of thrombolytic or beta blocker therapy.

To clarify these somewhat conflicting findings, Mant and Hicks systematically reviewed the literature to estimate effect sizes for therapies proven effective for AMI patients, based on clinical trials and meta-analyses (i.e., beta blockade, aspirin, fibrinolysis, and angiotensin converting enzyme inhibitors). Using these estimates and the proportion of patients eligible for treatment, the authors simulated the number of patients required to detect differences in care using either a "perfect system" for risk-adjusted mortality or a process-based quality of care audit. Plausible differences in lives lost were detectable with one year of data collection on mortality, consistent with AHRQ's recommendation to use a one-year minimum data collection period for IQI 15.

Relationship with 30-day mortality

Three studies compared adjusted in-hospital and 30-day mortality for AMI. Kristofferson et al., examined Norwegian hospital data (n=55 hospitals and n = 48,048 AMI patients) from 1997-2001 and found the adjusted mortality measures for in-hospital and 30-day AMI mortality to be highly correlated ($0.82 \leq r \leq 0.94$).[16] Borzecki et al. used AHRQ IQI (version 3.1) software and Veterans Health data from 2004-2007, for 119 facilities, to compare in-hospital and 30-day mortality rates for AMI.[17] They also found a strong correlation ($r \geq 70$, $p < 0.05$) between in-hospital and 30-day mortality, but the observed 30-day mortality rate (11.1%) for AMI was significantly higher than in-hospital mortality (7.2%). A third study by Drye et al. used Medicare claims data for admissions to non-federal acute care hospitals to investigate differences between in-hospital and 30-day mortality rates.[18] They found that the mean absolute risk-standardized difference between 30-day and in-hospital mortality was 6.0% for AMI. Moreover, of 3,135 hospitals studied, 8.2% had differing performance classifications for in-hospital and 30-day mortality. An increase in length-of-stay of one day was associated with an estimated absolute increase of 0.33% in risk-standardized AMI mortality.

Geographic and temporal variation

Studies have also examined geographic and temporal variation in IQI 15 rates. Kolte et al. used the 2003-2010 NIS to quantify regional differences in ST-elevation myocardial infarction treatment and in-hospital mortality.[19] They found risk-adjusted mortality to be higher in the Midwest (Odds Ratio [OR] 1.07; 95% Confidence Interval [CI] 1.05 to 1.09; p<0.001), South (OR 1.03; 95% C 1.01 to 1.05; p=0.001), and West (OR 1.06; 95% CI 1.04 to 1.08; p<0.001)), compared to the Northeast. Another study by Menees and colleagues addressed trends in door-to-balloon time for patients undergoing percutaneous coronary interventions (PCI) against in-hospital mortality using data on 96,738 patients admitted for ST-elevated myocardial infarction from the CathPCI Registry from July 2005-June 2009.[20] They found a significant decline between July 2005 – June 2006 and July 2008 – June 2009 in door-to-balloon times, as measured both by median times (83 minutes vs. 67 minutes, p<0.001) and by the percentage of patients with door-to-balloon times of 90 minutes or less (59.7% vs. 83.1%, p<0.001). However, despite these changes, there was no significant change in risk-adjusted in-hospital mortality over this time period (5.0% vs. 4.7%, p=0.34).

Risk adjustment

Numerous studies have established the importance of risk adjustment for AMI patients. As a result, researchers have developed a number of risk adjustment models. Normand et al. developed and validated two models, one of which was based on conditions likely to be present on admission and therefore applicable to comparisons of hospital-based care.[21] The claims-based model included 25 comorbidities not related to treatment. Hypertension (18.3%), diabetes (13.8%), and pulmonary disease (11.2%) were the most frequent

comorbidities in an AMI Medicare cohort of 164,427 patients. Frequently occurring comorbidities that were considered possibly related to hospital treatment, and therefore omitted, included congestive heart failure (33.9%), chronic angina (27.4%), and arrhythmias (25.2%). The same team developed another model using the clinical predictors available from the Cooperative Cardiovascular Project. From these and numerous other studies, the most important predictors of short-term AMI mortality have been shown to include age, previous AMI, tachycardia, pulmonary edema and other signs of congestive heart failure, hypotension and cardiogenic shock, anterior wall and Q-wave infarction, cardiac arrest, and serum creatinine or urea nitrogen.

Krumholz et al. compared seven models including a newly developed 7-variable clinical/demographic risk adjustment model for 30-day mortality in AMI patients.[22] The models based on clinical data demonstrated modestly better discrimination and calibration than two models based on ICD-9-CM codes (area under the receiver operating curve 0.74-0.78 versus 0.70-0.71, respectively). In addition, the clinical models classified hospital performance somewhat differently than the models based on administrative data. Such differences were further explored by Iezzoni et al., who used several proprietary products to estimate risk-adjusted AMI mortality, and found 40-60% disagreement in identifying the 10 best and 10 worst hospitals in a nationwide sample.[23,24] Adding full clinical data to administrative data for risk-adjustment, Pine found that 73% of Cleveland hospitals' expected mortality rates changed by less than one standard deviation, and none changed by more than two.[25] In St. Louis, 95% of hospitals' expected mortality rates changed by less than 0.5 standard deviations, and none changed by more than one. These estimates were better than those for other major medical conditions, including pneumonia, stroke, and congestive heart failure.[26] In the California Hospital Outcomes Project, the addition of clinical risk factors to a re-estimated model based on re-abstracted ICD-9-CM codes had a minimal effect on the difference in risk-adjusted mortality between low-mortality and high-mortality hospitals, although individual hospitals were affected.[27] In summary, these studies found that the method of risk-adjustment does affect which specific hospitals are identified as mortality outliers, but that the correlations within pairs of risk-adjusted or expected mortality rates are generally high (e.g., 0>0.80)[13] to 0.94,[26] and higher for AMI than for other medical conditions.

When risk adjustment models include ICD-9-CM conditions that may represent consequences of poor care, then discrimination is exaggerated.[22] Romano and Chan compared an administrative data set to a re-abstraction of diagnoses present on admission (POA), with two versions of the All Patient Refined-Diagnosis-Related Groups (APR-DRG), Risk of Mortality (ROM) and Severity of Illness (SOI).[28] The authors showed empirically that APR-DRGs predicted in-hospital mortality better when all diagnoses were included than when only POA diagnoses were included. Hospitals' expected mortality rates based on all re-abstracted ICD-9-CM codes were moderately correlated (r=0.72-0.77) with expected mortality rates based only on POA diagnoses. However, 2 of the 3 hospitals classified as having higher than expected mortality, 8 of the 23 hospitals classified as having neither higher nor lower than expected mortality, and 0 of the 4 hospitals classified as having lower than expected mortality, switched categories when diagnoses not present on admission were excluded from risk-adjustment. IQI 15 risk adjustment therefore only considers POA secondary diagnoses.

Scientific Acceptability

Metcalfe et al., searched Ovid Medline from 1950-2010 to identify studies (26 articles) that validated ICD-9, ICD-9-CM, and ICD-10 AMI case definitions. They then applied these ICD-based definitions to an ICD-9-CM/ICD-10 dual-coded dataset (n=4,008 inpatient records) to assess the criterion validity of the definitions (using chart review as the criterion standard) and to assess the impact of varied AMI case definitions on AMI in-hospital mortality estimates.[29] They found that AMI code 410 (ICD-9-CM) had the highest sensitivity (94%) and specificity (99%). The use of codes 410 (ICD-9-CM) and I21-I22 (ICD-10) to define AMI had high sensitivity (83.3%, 82.8% for ICD-9-CM and ICD-10, respectively) and positive predictive value (82.8%, 82.2%). The

percentage of in-hospital AMI mortality identified in the chart review, using these specific case definitions, was 7.6% (ICD-9-CM) and 6.6% (ICD-10).

Another study suggested that patients transferred to a second hospital may be counted twice for one episode of AMI (if these patients are not excluded, as in IQI 15), and sought to describe the impact of such double counting and transfer bias on the estimation of incidence rates and outcomes of AMI in the United States. Analyzing hospital discharge data from eight states, the study estimated that double count rates range from 10% to 15% for all states and increased over the 3 years. Moderate sized rural counties had the highest estimated double count rates at 15% to 20% with a few counties having estimated double count rates a high as 35% to 50%. Older patients and females were less likely to be double counted ($p < 0.05$).[30] This problem is addressed in the design of IQI 15 by excluding inter-hospital transfers from the transferring facility, and adjusting for their marginal risk at the receiving facility.

References

1. Allison JJ, Kiefe CI, Weissman NW, et al. Relationship of hospital teaching status with quality of care and mortality for Medicare patients with acute MI. *JAMA.* 2000; 284(10):1256-1262.
2. Polanczyk CA, Lane A, Coburn M, Philbin EF, Dec GW, DiSalvo TG. Hospital outcomes in major teaching, minor teaching, and nonteaching hospitals in New York state. *The American Journal of Medicine.* 2002; 112(4):255-261.
3. Jena AB, Sun EC, Romley JA. Mortality among high-risk patients with acute myocardial infarction admitted to U.S. teaching-intensive hospitals in July: a retrospective observational study. *Circulation.* 2013;128(25):2754-2763.
4. Patel MS, Volpp KG, Small DS, et al. Association of the 2011 ACGME resident duty hour reforms with mortality and readmissions among hospitalized Medicare patients. *JAMA.* 2014;312(22):2364-2373.
5. Carretta HJ, Chukmaitov A, Tang A, Shin J. Examination of hospital characteristics and patient quality outcomes using four inpatient quality indicators and 30-day all-cause mortality. *American Journal of Medical Quality : the official Journal of the American College of Medical Quality.* 2013;28(1):46-55.
6. Chen J, Rathore SS, Radford MJ, Krumholz HM. JCAHO accreditation and quality of care for acute myocardial infarction. *Health affairs (Project Hope).* 2003;22(2):243-254.
7. Vakili BA, Kaplan R, Brown DL. Volume-outcome relation for physicians and hospitals performing angioplasty for acute myocardial infarction in New York state. *Circulation.* 2001;104(18):2171-2176.
8. Canto JG, Every NR, Magid DJ, et al. The volume of primary angioplasty procedures and survival after acute myocardial infarction. National Registry of Myocardial Infarction 2 Investigators. *The New England Journal of Medicine.* 2000;342(21):1573-1580.
9. Canto JG, Rogers WJ, French WJ, Gore JM, Chandra NC, Barron HV. Payer status and the utilization of hospital resources in acute myocardial infarction: a report from the National Registry of Myocardial Infarction 2. *Archives of Internal Medicine.* 2000;160(6):817-823.
10. Chen EW, Canto JG, Parsons LS, et al. Relation between hospital intra-aortic balloon counterpulsation volume and mortality in acute myocardial infarction complicated by cardiogenic shock. *Circulation.* 2003;108(8):951-957.
11. Every NR, Maynard C, Schulman K, Ritchie JL. The association between institutional primary angioplasty procedure volume and outcome in elderly Americans. *The Journal of Invasive Cardiology.* 2000;12(6):303-308.
12. Magid DJ, Calonge BN, Rumsfeld JS, et al. Relation between hospital primary angioplasty volume and mortality for patients with acute MI treated with primary angioplasty vs thrombolytic therapy. *JAMA.* 2000;284(24):3131-3138.
13. Krumholz HM, Rathore SS, Chen J, Wang Y, Radford MJ. Evaluation of a consumer-oriented internet health care report card: the risk of quality ratings based on mortality data. *JAMA.* 2002;287(10):1277-1287.
14. McCrum ML, Joynt KE, Orav EJ, Gawande AA, Jha AK. Mortality for publicly reported conditions and overall hospital mortality rates. *JAMA Internal Medicine.* 2013;173(14):1351-1357.

15. Romley JA, Jena AB, O'Leary JF, Goldman DP. Spending and mortality in US acute care hospitals. *The American Journal of Managed Care.* 2013;19(2):e46-54.
16. Kristoffersen DT, Helgeland J, Clench-Aas J, Laake P, Veierod MB. Comparing hospital mortality--how to count does matter for patients hospitalized for acute myocardial infarction (AMI), stroke and hip fracture. *BMC Health Services Research.* 2012;12:364.
17. Borzecki AM, Christiansen CL, Chew P, Loveland S, Rosen AK. Comparison of in-hospital versus 30-day mortality assessments for selected medical conditions. *Medical Care.* 2010;48(12):1117-1121.
18. Drye EE, Normand SL, Wang Y, et al. Comparison of hospital risk-standardized mortality rates calculated by using in-hospital and 30-day models: an observational study with implications for hospital profiling. *Annals of Internal Medicine.* 2012;156(1 Pt 1):19-26.
19. Kolte D, Khera S, Aronow WS, et al. Regional variation across the United States in management and outcomes of ST-elevation myocardial infarction: analysis of the 2003 to 2010 nationwide inpatient sample database. *Clinical Cardiology.* 2014;37(4):204-212.
20. Menees DS, Peterson ED, Wang Y, et al. Door-to-balloon time and mortality among patients undergoing primary PCI. *The New England Journal of Medicine.* 2013;369(10):901-909.
21. Normand ST, Glickman ME, Sharma RG, McNeil BJ. Using admission characteristics to predict short-term mortality from myocardial infarction in elderly patients. Results from the Cooperative Cardiovascular Project. *JAMA.* 1996;275(17):1322-1328.
22. Krumholz HM, Chen J, Wang Y, Radford MJ, Chen YT, Marciniak TA. Comparing AMI mortality among hospitals in patients 65 years of age and older: evaluating methods of risk adjustment. *Circulation.* 1999;99(23):2986-2992.
23. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Predicting who dies depends on how severity is measured: implications for evaluating patient outcomes. *Annals of Internal Medicine.* 1995;123(10):763-770.
24. Landon B, Iezzoni LI, Ash AS, et al. Judging hospitals by severity-adjusted mortality rates: the case of CABG surgery. *Inquiry : a Journal of Medical Care Organization, Provision and Financing.* 1996;33(2):155-166.
25. Pine M, Norusis M, Jones B, Rosenthal GE. Predictions of hospital mortality rates: a comparison of data sources. *Annals of Internal Medicine.* 1997;126(5):347-354.
26. Pine M, Jones B, Lou YB. Laboratory values improve predictions of hospital mortality. *International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua.* 1998;10(6):491-501.
27. *Second Report of the California Hospitals Outcomes Project, May 1996, Acute Myocardial Infarction.* Sacramento, CA 1996.
28. Romano PS, Chan BK. Risk-adjusting acute myocardial infarction mortality: are APR-DRGs the right tool? *Health Services Research.* 2000;34(7):1469-1489.
29. Metcalfe A, Neudam A, Forde S, et al. Case definitions for acute myocardial infarction in administrative databases and their impact on in-hospital mortality rates. *Health Services Research.* 2013;48(1):290-318.
30. Westfall JM, McGloin J. Impact of double counting and transfer bias on estimated rates and outcomes of acute myocardial infarction. *Medical Care.* 2001;39(5):459-468.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
AHRQ_IQI15_NQF_0730_Measure_Evidence_Form.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Acute Myocardial Infarction (AMI) is an emergent condition for which significant morbidity and mortality may result from delayed, inappropriate or low quality treatment. Processes that have been identified to lower AMI mortality include: timely electrocardiography (ECG), early percutaneous coronary intervention (PCI) for ST-elevation MI, or fibrinolytic therapy if PCI is not available, aspirin administration at arrival, afterload reduction with renin-angiotensin system blockade among patients with impaired left ventricular function, and aspirin/statin prescription initiation during hospitalization with continuation at discharge. The 30-day risk-standardized AMI mortality rate is currently a CMS quality metric. In-hospital risk-adjusted mortality rates allow users without access to post-hospitalization death data, to assess and compare AMI outcomes.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This table is also included in the supplemental files.

Table 1. Reference Population Rate and Distribution of Hospital Performance of IQI 15 Acute Myocardial Infarction (AMI) Mortality Rate

Overall Reference Population Rate

Year	Number Hospitals (Numerator)1	Population at Risk (Denominator)1	Outcome of Interest Observed Rate Per 10001	
2012	2,978	24,890	441,557	56.37
2011	2,835	23,995	411,209	58.35
20103	4,063	30,917	512,422	60.34
20093	4,023	31,543	503,971	66.41
20083	4,029	34,093	513,338	68.94

Distribution of Hospital-level Observed Rates in Reference Population

Year	Number of Hospitals	Distribution of Observed Hospital-level Rates per 1000 (p=percentile)2						
		Mean	SD	p5	p25	Median	p75	p95
2012	2,978	104.33	150.90	0.00	33.90	58.82	115.38	375.00
2011	2,835	111.90	153.88	0.00	36.91	63.45	125.00	400.00

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2008-2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 4.5 and 5.0)
1The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined population of all hospitals included in the reference population data (denominator).
2The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals included in the dataset, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile.
32008-2010 data are calculated using Version 4.5 of the QI Software and all states included in the SID for those years. Version 4.5 includes a "prediction module" which is used to account for missing present on admission flags. In Version 5.0, the "prediction module" has been removed and the reference population is limited to states and hospitals with present on admission data. These differences may lead to some discontinuity in the observed rates between 2010 and 2011, since many states did not report POA data prior to 2011. The number of states reporting consistent POA has increased from 2008-2012.

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.
Not applicable

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

This table is also included in the supplemental files.

Table 2. Observed AMI Mortality Rates per 1,000 (IQI 15), by patient and hospital characteristics, 2012

Patient Characteristics	Estimate	Std Error	p-value	(Ref Grp = *)
Total U.S.	56.296	0.351		
Patient Characteristics				
Age Groups:				
18-44*	18.064	0.878		
45-64	29.323	0.419	0.000	
65 and over	77.626	0.539	0.000	
Gender:				
Male*	50.623	0.427		
Female	65.174	0.602	0.000	
Patient Zip Code Median Income				
First quartile (lowest income)	57.589	0.658	0.237	
Second quartile	55.447	0.681	0.916	
Third quartile	55.151	0.712	0.949	
Fourth quartile (highest income)*	56.864	0.771		
Location of patient residence (NCHS):				
Large central metropolitan	58.758	0.695	0.000	
Large fringe metropolitan*	54.548	0.703		
Medium metropolitan	55.131	0.793	0.291	
Small metropolitan	54.810	1.082	0.420	
Micropolitan	58.600	1.051	0.001	
Noncore	54.815	1.206	0.424	
Expected payment source:				
Private insurance*	27.038	0.488		
Medicare	74.327	0.525	0.000	
Medicaid	41.888	1.205	0.000	
Other insurance	40.254	1.731	0.000	
Uninsured / self-pay / no charge	35.226	1.038	0.000	
Hospital Characteristic:				
Location of Care:				
Northeast*	61.505	0.804		
Midwest	52.154	0.762	1.000	
South	54.161	0.550	1.000	
West	59.138	0.796	0.982	

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 5.0.1)
*Reference for p-value test statistics.
NCHS - National Center for Health Statistics designation for urban-rural locations.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Not applicable

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Table 1 in section 1.b.2 shows that the total U.S. observed rate for AMI mortality in 2012 was 56.4 per 1000, representing an estimated total of 24,890 deaths. From 2008-2012, rates decreased from 68.9 in 2008 to 56.4 in 2012. Note that between 2010 and 2011, the reference population changed substantially, notably limiting to states with strong POA data and to community hospitals (i.e., excluding children's hospitals, rehabilitation hospitals, etc.).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2008-2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 4.5 and 5.0)

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable

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 subcriteria 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 : Acute Myocardial Infarction

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

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.)*

http://www.qualityindicators.ahrq.gov/Modules/iqi_resources.aspx

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** Technical_Specs_IQI15_v5.0.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.

As standard protocol, the AHRQ QI program annually updates all measures with Fiscal Year coding changes, refinements based on stakeholder input, refinements to improve specificity and sensitivity based on additional analyses, and necessary software changes. In addition, approximately every two years, AHRQ updates the risk adjustment parameter estimates and composite weights based

on the most recent year of data (i.e., the most current reference population possible). The refined measures are tested and confirmed to be valid and reliable prior to release of the updated software.

Since the last endorsement (version 4.3), no changes have been made to the specification of IQI 15.

In Version 5.0 (released April 2015), the reference population has been updated to limit the population to states with strong present on admission (POA) data and community hospitals. The software no longer supports the "prediction module" for POA imputation, and instead requires user supplied POA data.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of in-hospital deaths among cases meeting the inclusion and exclusion rules for the denominator.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The time period is one year for users with a complete sample of hospital discharges (i.e., "all payer" data). Note that the signal variance parameters assume a one-year time period. Users may use longer time periods if desired.

S.6. **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, 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.
Number of deaths (DISP=20 in AHRQ's Healthcare Cost and Utilization Project datasets) among cases meeting the inclusion and exclusion rules for the denominator.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Discharges, for patients ages 18 years and older, with a principal ICD-9-CM diagnosis code for AMI.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
ICD-9-CM AMI diagnosis codes (initial or unspecified episode of care):

41000 AMI ANTEROLATERAL, UNSPEC
41001 AMI ANTEROLATERAL, INIT
41010 AMI ANTERIOR WALL, UNSPEC
41011 AMI ANTERIOR WALL, INIT
41020 AMI INFEROLATERAL, UNSPEC
41021 AMI INFEROLATERAL, INIT
41030 AMI INFEROPOST, UNSPEC
41031 AMI INFEROPOST, INITIAL
41040 AMI INFERIOR WALL, UNSPEC
41041 AMI INFERIOR WALL, INIT
41050 AMI LATERAL NEC, UNSPEC
41051 AMI LATERAL NEC, INITIAL
41060 TRUE POST INFARCT, UNSPEC
41061 TRUE POST INFARCT, INIT
41070 SUBENDO INFARCT, UNSPEC
41071 SUBENDO INFARCT, INITIAL
41080 AMI NEC, UNSPECIFIED
41081 AMI NEC, INITIAL

41090 AMI NOS, UNSPECIFIED
41091 AMI NOS, INITIAL

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclude cases:

- transferred to another short-term hospital, for whom the outcome at hospital discharge was unknown
- admitted for treatment of pregnancy, childbirth, and puerperium
- with missing discharge disposition, gender, age, quarter, year, or principal diagnosis

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Exclude cases:

- transferred to another short-term hospital (DISP=2)
- with Major Diagnosis Category (MDC) 14 (pregnancy, childbirth, and puerperium)
- with missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1=missing)

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age (in 5-year age groups), All Patient Refined Diagnosis Related Groups (APR DRGs) with Risk of Mortality (ROM) scores, Major Diagnosis Categories (MDC) based on the principal diagnosis, and transfer in from another acute care hospital. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., hospital). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.

The specific covariates for this measure are as follows:

Parameter	Label
Age	18 to 39
Age	40 to 44
Age	45 to 49
Age	50 to 54
Age	55 to 59
Age	65 to 79
Age	80 to 84
Age	85+
APR-DRG	161-(1-2) CARDIAC DEFIBRILLATOR & HEART ASSIST IMPLANT, Risk of mortality (ROM) 1 - 2
APR-DRG	161-(3-4) CARDIAC DEFIBRILLATOR & HEART ASSIST IMPLANT, Risk of mortality (ROM) 3 - 4
APR-DRG	162-(1,2)CARDIAC VALVE PROCEDURES W CARDIAC CATHETERIZATION, ROM 1 and 2
APR-DRG	162-3 CARDIAC VALVE PROCEDURES W CARDIAC CATHETERIZATION, ROM 3
APR-DRG	162-4 CARDIAC VALVE PROCEDURES W CARDIAC CATHETERIZATION, ROM 4
APR-DRG	165-(1,2) CORONARY BYPASS W CARDIAC CATH OR PERCUTANEOUS CARDIAC PROC, ROM 1 and 2
APR-DRG	165-3 CORONARY BYPASS W CARDIAC CATH OR PERCUTANEOUS CARDIAC PROC, ROM 3
APR-DRG	165-4 CORONARY BYPASS W CARDIAC CATH OR PERCUTANEOUS CARDIAC PROC, ROM 4
APR-DRG	173-(1-4) OTHER VASCULAR PROCEDURES, ROM 1-4

APR-DRG 174-2 PERCUTANEOUS CARDIOVASCULAR PROCEDURES W AMI, ROM 2
APR-DRG 174-3 PERCUTANEOUS CARDIOVASCULAR PROCEDURES W AMI, ROM 3
APR-DRG 174-4 PERCUTANEOUS CARDIOVASCULAR PROCEDURES W AMI, ROM 4
APR-DRG 190-1 ACUTE MYOCARDIAL INFARCTION, ROM 1
APR-DRG 190-2 ACUTE MYOCARDIAL INFARCTION, ROM 2
APR-DRG 190-3 ACUTE MYOCARDIAL INFARCTION, ROM 3
APR-DRG 190-4 ACUTE MYOCARDIAL INFARCTION, ROM 4
MDC 5 CIRCULATORY SYSTEM, DISEASES & DISORDERS
TRNSFER TRANSFER IN FROM ANOTHER ACUTE CARE HOSP (If ASOURCE='2' (Another Hospital) or POINTOFORIGINUB04=´4´ (Transfer from a Hospital), then TRNSFER=1)

Source: http://qualityindicators.ahrq.gov/Downloads/Modules/IQI/V50/Parameter_Estimates_IQI_50.pdf.pdf

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
The observed rate is the number of discharge records where the patient experienced the QI adverse event divided by the number of discharge records at risk for the event. The expected rate is a comparative rate that incorporates information about a reference population that is not part of the user's input dataset – what rate would be observed if the expected level of care observed in the reference population and estimated with risk adjustment regression models, were applied to the mix of patients with demographic and comorbidity distributions observed in the user's dataset? The expected rate is calculated only for risk-adjusted indicators.

The expected rate is estimated for each person using a generalized estimating equations (GEE) approach to account for correlation at the hospital or provider level.

The risk-adjusted rate is a comparative rate that also incorporates information about a reference population that is not part of the input dataset – what rate would be observed if the level of care observed in the user's dataset were applied to a mix of patients with demographics and comorbidities distributed like the reference population? The risk adjusted rate is calculated using the indirect method as observed rate divided by expected rate multiplied by the reference population rate. The smoothed rate is the weighted average of the risk-adjusted rate from the user's input dataset and the rate observed in the reference population; the smoothed rate is calculated with a shrinkage estimator to result in a rate near that from the user's dataset if the provider's rate is estimated in a stable fashion with minimal noise, or to result in a rate near that of the reference population if the variance of the estimated rate from the input dataset is large compared with the hospital-to-hospital variance estimated from the reference population. Thus, the smoothed rate is a weighted average of the risk-adjusted rate and the reference population rate, where the weight is the signal-to-noise ratio. In practice, the smoothed rate brings rates toward the mean, and tends to do this more so for outliers (such as rural hospitals).

For additional information, please see supporting information in the Quality Indicator Empirical Methods.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation*

Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Exclude cases with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing). Missingness on these variables, in aggregate, almost never exceeds 1% of eligible records.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
While the measure is tested and specified using data from the Healthcare Cost and Utilization Project (HCUP) (see section 1.1 and 1.2 of the measure testing form), the measure specifications and software are specified to be used with any ICD-9-CM-coded administrative billing/claims/discharge dataset with Present on Admission (POA) information. Note that in Version 5.0, the AHRQ QI software no longer supports prediction of POA status using an embedded prediction module. Users are expected to provide POA data.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
AHRQ_IQI15_Measure_Testing_Form.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0730
Measure Title: Acute Myocardial Infarction (AMI) Mortality Rate
Date of Submission: 6/30/2015
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/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*).

All analyses were completed using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), 2008-2012. HCUP is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ).[1] HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. The HCUP SID contain the universe of the inpatient discharge abstracts in participating States, translated into a uniform format to facilitate multi-State comparisons and analyses. All states provide data for community hospitals and together, the SID encompasses about 97 percent of all U.S. community hospital discharges. For the analyses presented here, we use 36 of the 45 states that participated in 2012, and 82 percent of the U.S. community hospital discharges, for a total of about 30 million hospital discharges from community hospitals. As defined by the American Hospital Association, community hospitals are all non-Federal, short-term, general or other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are public and academic medical centers, specialty hospitals such as obstetrics–gynecology, ear–nose–throat, orthopedic and pediatric institutions. Short-stay rehabilitation, long-term acute care hospitals are excluded from the data used for the reported analyses.

Each of the 36 states included in the dataset report information about whether a diagnosis was present on admission (POA) and information on the timing of procedures during the hospitalization. POA data[1] is important to distinguish complications that occur in-hospital from diagnoses that existed prior to hospitalization. For all PSIs, the POA flag is used to exclude cases from the numerator when the condition of interest is present on admission and to exclude complications that occur in-hospital from risk adjustment. Edit

[1] Present-on -Admission was added as a data element to the uniform bill form (UB-04) effective October 1, 2007, and hospitals incurred a payment penalty for not including POA on Medicare records beginning October 1, 2008. Each of the several diagnoses in a discharge record can be flagged as "present at the time the order for inpatient admission occurs" or not (see http://www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm).

checks on POA were developed using a separate analysis of HCUP databases that examined POA coding in the 2011 SID at hospitals that were required to report POA to CMS. The edits identify general patterns of suspect reporting of POA. The edits do not evaluate whether a valid POA value (e.g., Y or N) is appropriate for the specific diagnosis. There are three hospital-level edit checks:

1. Indication that a hospital has POA reported as Y on all diagnoses on all discharges
2. Indication that a hospital has POA reported as missing on all non-Medicare discharges
3. Indication that a hospital reported POA as missing on all nonexempt diagnoses for 15 percent or more of discharges. The cut-point of 15 percent was determined by 2 times the standard deviation plus the mean of the percentage for hospitals required to report POA to CMS.

Hospitals that failed any of the edit checks were excluded from the dataset.

The SID data elements include ICD-9-CM coded principal and secondary diagnoses and procedures, additional detailed clinical and service information based on revenue codes, admission and discharge status, patient demographics, expected payment source (Medicare, Medicaid, private insurance as well as the uninsured), total charges and length of stay (www.hcup-us.ahrq.gov).

[1]HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2008-2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 5.0)

1.3. What are the dates of the data used in testing? 2008-2012

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.26***)	**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*)

Table 1. Reference Population Rate and Distribution of Hospital Performance IQI 15 Acute Myocardial Infarction (AMI) Mortality Rate

Overall Reference Population Rate				
Year	**Number**	**Outcome of Interest**	**Population at Risk**	**Observed Rate**

	Hospitals	(Numerator)[1]	(Denominator)[1]	Per 1000[1]
2012	2,978	24,890	441,557	56.37
2011	2,835	23,995	411,209	58.35
2010[3]	4,063	30,917	512,422	60.34
2009[3]	4,023	31,543	503,971	66.41
2008[3]	4,029	34,093	513,338	68.94

Distribution of Hospital-level Observed Rates in Reference Population

Year	Number of Hospitals	Distribution of Observed Hospital-level Rates per 1000 (p=percentile)[2]						
		Mean	SD	p5	p25	Median	p75	p95
2012	2,978	104.33	150.90	0.00	33.90	58.82	115.38	375.00
2011	2,835	111.90	153.88	0.00	36.91	63.45	125.00	400.00

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2008-2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 4.5 and 5.0)

[1]The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined population of all hospitals included in the reference population data (denominator).

[2]The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals included in the dataset, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile.

[3]2008-2010 data are calculated using Version 4.5 of the QI Software and all states included in the SID for those years. Version 4.5 includes a "prediction module" which is used to account for missing present on admission flags. In Version 5.0, the "prediction module" has been removed and the reference population is limited to states and hospitals with present on admission data. These differences may lead to some discontinuity in the observed rates between 2010 and 2011, since many states did not report POA data prior to 2011. The number of states reporting consistent POA has increased from 2008-2012.

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

See 1.5 (Table 1)

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.

Some tests require comparisons of two years of data (2011-2012). When no comparisons are required for the test, only 2012 data are used. For the sake of providing 5-year comparisons, we also provide rates from 2008-2010 using Version 4.5 of the software. Version 5.0 could not be calculated for 2008-2010 due to missing POA data. POA data is used in the risk adjustment model.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when

SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Age and sex were the only patient-level sociodemographic variables that were available and analyzed in the data used for measure development and testing. The development data sets generally include race/ethnicity, principal expected source of payment, and zip code of residence, which could be used to capture socioeconomic characteristics at an ecological (community) level. However, these variables were not used in our analyses, based on our conceptual description (logical rationale or theory informed by literature and content experts) of the causal pathway between these factors, patient clinical factors, quality of care, and outcome, described in Section 2b4.3 below.

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*)

Signal-to-noise. The signal-to-noise ratio refers to the entire population of US hospitals, comparing the degree to which rates are different from hospital to hospital (the signal) to how stable the rates are within hospitals (the noise). This metric is a stringent measure of reliability that takes into account the observed distribution of rates within a reference population. An indicator with a low signal-to-noise ratio may not be able to distinguish differences in quality between hospitals, or may identify differences inconsistently within the same time period. An indicator with a high signal-to-noise ratio will be more likely to consistently distinguish performance differences between hospitals (e.g. one hospital performs better than others).

The signal-to-noise ratio is estimated for each hospital. The overall signal-to-noise estimate is an average of hospital-level signal to noise ratios weighted by hospital size. Hospital size is calculated as the number of eligible discharges for IQI15. Weighting by hospital size reduces the impact of hospitals that have very small denominators (the number of patients at risk). Small hospitals admit very few patients at risk for AMI mortality.

Because the signal-to-noise ratio quantifies the ability to consistently discriminate one hospital's performance from the other hospitals in the population, it is sensitive to the distribution of hospital sizes as well as the distribution of observed rates in the reference population. If the hospitals in a population all have performance in a narrow range, it is more difficult to reliably distinguish between hospitals' performance than when hospital performance is spread out over a much wider range. For example, if all hospitals have nearly perfect performance, it will be impossible to distinguish between them. As a consequence, if the distribution of hospital rates changes over time, the signal-to-noise ratio will also change.

There is no universally accepted threshold of "adequate" signal to noise ratio. Different methods of calculating reliability and signal-to-noise result in different distributions of reliability scores. In addition, "adequate" depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8. To account for the uncertainty (noise) in a hospital's performance due to reliability concerns stemming from low volume, smoothed rates can be calculated.

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

Table 2. Signal-to-Noise Ratio by Size Decile IQI 15 Acute Myocardial Infarction (AMI) Mortality

Size Decile	Number of Hospitals	Avg. Number of Discharges per Hospital in Decile	Avg. Signal-to-Noise Ratio for Hospitals in Decile
1	266	4.2	0.16937
2	266	8.2	0.24352
3	267	14.9	0.3386
4	266	27.7	0.47658
5	267	51.5	0.58262
6	266	100.8	0.66445
7	267	166.2	0.73103
8	266	249.1	0.78999
9	267	361.3	0.83704
10	266	672.2	0.89988
Overall	2664	165.6	0.74937

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 5.0)

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?*)

Signal to noise ratios were smaller for hospitals with fewer than 15 qualifying discharges per year (average signal-to-noise ratio less than 0.34). Smoothed rates, which are recommended for all hospitals (and are implemented in the AHRQ software), address reliability concerns particularly for small hospitals. Hospitals with more than 28 discharges on average have risk adjusted rates with moderate to high reliability (average signal-to-noise ratio of 0.48 to 0.90). Overall, the signal to noise ratio for this indicator is strong with an overall signal-to-noise ratio of 0.75.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)

Critical Data Elements

The most critical data element for IQI 15 is the principal diagnosis field, which is used to identify patients admitted principally for treatment of AMI. Several published studies have reported on the sensitivity, specificity, and/or positive predictive value (PPV) of hospital administrative data for the purpose of identifying patients with a principal diagnosis of AMI (ICD-9-CM 410.x0 or 410.x1):

1: Metcalfe A, Neudam A, Forde S, Liu M, Drosler S, Quan H, Jetté N. Case definitions for acute myocardial infarction in administrative databases and their impact on in-hospital mortality rates. Health Serv Res. 2013 Feb;48(1):290-318.
2: Meehan TP, Hennen J, Radford MJ, Petrillo MK, Elstein P, Ballard DJ. Process and outcome of care for acute myocardial infarction among Medicare beneficiaries in Connecticut: a quality improvement demonstration project. Ann Intern Med. 1995 Jun 15;122(12):928-36.
3: Choma NN, Griffin MR, Huang RL, Mitchel EF Jr, Kaltenbach LA, Gideon P, Stratton SM, Roumie CL. An algorithm to identify incident myocardial infarction using Medicaid data. Pharmacoepidemiol Drug Saf. 2009 Nov;18(11):1064-71.
4: Kiyota Y, Schneeweiss S, Glynn RJ, Cannuscio CC, Avorn J, Solomon DH. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. Am Heart J. 2004 Jul;148(1):99-104.
5: Petersen LA, Wright S, Normand SL, Daley J. Positive predictive value of the diagnosis of acute myocardial infarction in an administrative database. J Gen Intern Med. 1999 Sep;14(9):555-8.
6: Cutrona SL, Toh S, Iyer A, Foy S, Daniel GW, Nair VP, Ng D, Butler MG, Boudreau D, Forrow S, Goldberg R, Gore J, McManus D, Racoosin JA, Gurwitz JH. Validation of acute myocardial infarction in the Food and Drug Administration's Mini-Sentinel program. Pharmacoepidemiol Drug Saf. 2013 Jan;22(1):40-54.
7: Yeh RW, Sidney S, Chandra M, Sorel M, Selby JV, Go AS. Population trends in the incidence and outcomes of acute myocardial infarction. N Engl J Med. 2010 Jun 10;362(23):2155-65.
8: Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, Hsia DC. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 1992 Feb;82(2):243-8.

Empirical Validity

We tested the empirical validity of IQI 15 in a variety of ways:

1. Previous studies have demonstrated a relatively high correlation between inpatient and 30-day mortality for AMI, with a moderate kappa value. We empirically compared CMS 30-day risk-standardized mortality rates (RSMR) for AMI to IQI 15 rates (restricted to hospitalizations meeting the CMS measure's denominator criteria) by estimating Spearman rank correlations.
2. We evaluated the association between process measures involving AMI care and IQI 15 rates with the expectation that the two should generally be inversely related. The process measures included: (1) Median time to ECG (OP-5); (2) Fibrinolytic therapy received within 30 minutes of emergency department arrival (OP-2); (3) Fibrinolytic therapy received within 30 minutes of hospital arrival (AMI-7a); (4) Timing of receipt of primary percutaneous coronary intervention (AMI-8a); (5) Aspirin prescribed at discharge (AMI-2); and (6) Statin prescribed at discharge (AMI-10). We estimated Spearman rank correlations between these process measures and IQI 15 rates at the hospital level.
3. Teaching hospitals have been associated with improved outcomes from AMI, probably by virtue of closer adherence to current evidence-based guidelines. We determined the association between hospital teaching status and IQI 15 rates.
4. Assuming that hospitals that perform a high volume of percutaneous coronary intervention procedures are adept at management of patients with AMI, we assessed the relationship between

percutaneous coronary intervention (PCI) volume (as expressed by AHRQ's IQI 6) and IQI 15 rates. We tested for differences in rates across quartiles of PCI volume.

5. IQI 15 performance should be associated with performance on AHRQ's other IQIs that involve mortality from cardiovascular conditions or procedures. We assessed the relationship between IQIs 12 (Coronary Artery Bypass Graft Mortality Rate), 16 (Heart Failure Mortality Rate), and 30 (Percutaneous Coronary Intervention Mortality Rate) and IQI 15 rates. We estimated Spearman rank correlations among these measures.
6. Previous evidence suggests a meaningful volume-outcome association for AMI. For this reason, IQI 15 performance should be associated with volume of AMI hospitalizations. We assessed the relationship between AMI volume, as expressed by the IQI 15 denominator, and IQI 15 rates using the Spearman rank correlation coefficient.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Critical Data Elements

In a systematic review of alternative case definitions for AMI in administrative databases, Metcalfe et al. identified 8 ICD-9-CM based definitions that have been validated in peer-reviewed studies. When these 8 definitions were applied to a Canadian (Calgary, Alberta) database with 4,008 inpatient records from 2003 that had been independently coded in both ICD-9-CM and ICD-10, the codes used in the current IQI 15 definition (410.x0 and 410.x1) had the highest PPV (84.0%) and acceptable sensitivity of 81.1%. However, these authors did not focus on the principal diagnosis position, and did not limit their "gold standard" to patients admitted for treatment of AMIs that occurred in the community.

The following authors have validated this case definition (or a slightly broader case definition including all 410 codes) based only on the principal diagnosis field in US data:

1. Fisher et al., 1992: sensitivity = 94.0%, PPV = 92.0% in 1984-1985 data (using 410.x)
2. Meehan et al., 1995; PPV = 95.7% in 1988-1991 data (using 410.x)
3. Petersen et al., 1999: PPV = 96.9% in 1994-1995 data (also applied LOS ≥3 days)
4. Kiyota et al., 2004: PPV = 95.1% in 1999-2000 data
5. Choma et al., 2009: PPV = 92.8% in 1999-2004 data (using 410.x, also required LOS ≥3 days)
6. Yeh et al., 2010: PPV = 96.7% in 1999-2008 data
7. Cutrona et al., 2013: PPV =86.0% in 2009 data

Empirical Validity

The following analyses address the hospital-level Spearman rank correlation between IQI 15 risk-adjusted rates and process measure adherence. Unless otherwise specified, the first estimate is based on 2011 data and the second estimate is based on 2012 data.

1. CMS OP-2, Fibrinolytic therapy received within 30 minutes (NQF 0288) – R=-0.102 and -0.263 (i.e., hospitals with higher IQI15 mortality have less timely use of fibrinolytic therapy, or worse process performance, but only 763 hospitals were eligible for this analysis).
2. CMS OP-5, Median time to ECG (NQF 0289) – R=-0.062 and -0.051 (i.e., hospitals with higher IQI15 mortality have shorter median time to electrocardiogram (ECG), or better process performance, but correlations are weak).
3. AMI-2, Aspirin prescribed at discharge (NQF 0142) – R=-0.192 and -0.217 (i.e., hospitals with higher IQI15 mortality have less use of aspirin at discharge, or worse process performance).
4. AMI-7a, Fibrinolytic (thrombolytic) agent received within 30 minutes of hospital arrival (NQF 0164) – R=-0.018 (i.e., no significant correlation between IQI15 mortality and timely fibrinolytic therapy, but only 293 hospitals were eligible for this analysis).

5. AMI-8a, Receipt of primary percutaneous coronary intervention (PCI) within 90 minutes of hospital arrival (NQF 0163) – R=-0.207 and -0.238 (i.e., hospitals with higher IQI15 mortality have less timely use of PCI, or worse process performance).
6. AMI-10, Statin prescribed at discharge (NQF 0639) – R=-0.333 (i.e., hospitals with higher IQI15 mortality have less use of statin at discharge, or worse process performance).

With respect to hospital-level Spearman rank correlations between IQI 15 risk-adjusted rates and risk-standardized 30-day mortality rates for Medicare FFS patients (from the Yale-CMS model), analyses are limited to 13 states in the reference population for which Medicare FFS records could be identified:
1. Comparisons involving IQI 15 estimates based on ALL patients (i.e., all payers, all ages) are modest but statistically significant (r=0.161 and 0.144).
2. Comparisons involving IQI 15 estimates limited to Medicare FFS beneficiaries are similar to all patient-comparisons (r=0.147 and 0.139).

With respect to teaching status, teaching hospitals had significantly and substantially lower IQI 15 risk-adjusted mortality rates than non-teaching hospitals (i.e., weighted means 6.2% versus 12.4% in 2011, 6.0% versus 11.5% in 2012).

With respect to PCI volume, high-volume (top quartile by volume) PCI hospitals had significantly and substantially lower IQI 15 risk-adjusted mortality rates than low-volume (bottom quartile) PCI hospitals (i.e., weighted means 3.6% versus 7.0% in 2011, 3.5% versus 7.1% in 2012). Similarly, hospitals meeting the American College of Cardiology threshold for sufficient PCI volume (200 cases/year) had lower IQI 15 risk-adjusted mortality than PCI-performing hospitals with lower PCI volume (3.4% versus 4.2% in 2011, 3.3% versus 3.9% in 2012).

With respect to hospital-level Spearman rank correlations between IQI 15 risk-adjusted rates and risk-adjusted inpatient mortality rates for other cardiovascular diseases and procedures:
1. IQI 12 – Among the 913 hospitals that performed coronary artery bypass graft (CABG) surgery, Spearman rank correlations with CABG mortality were 0.222 (2011) and 0.226 (2012).
2. IQI 16 – Among the 2574 hospitals that treated both AMI and heart failure (HF), Spearman rank correlations with HF mortality were 0.172 (2011) and 0.130 (2012).
3. IQI 30 – Among the 1333 hospitals that performed PCI, Spearman rank correlations with PCI mortality were 0.406 (2011) and 0.407 (2012).

With respect to the volume-mortality relationship for acute inpatient AMI care, high-volume (top quartile by volume) AMI hospitals had significantly and substantially lower IQI 15 risk-adjusted rates than low-volume (bottom quartile) AMI hospitals (i.e., weighted means 3.8% versus 20.1% in 2011).

2b2.4. 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?*)

Most studies have estimated the accuracy of the current IQI 15 case definition in ICD-9-CM coded administrative data from US hospitals in the range of 92.8-96.9%; one study reported a lower estimate (PPV=86.0%; 95% CI 79.2-91.2%) but this study had relatively low inter-rater reliability for its gold standard of cardiologist review (kappa=0.60; 95% CI: 0.42, 0.78). Variation across hospitals was nonsignificant when tested.

Almost all of the expected construct validity relationships were supported. Specifically, hospitals with higher risk-adjusted inpatient mortality, according to IQI 15, also reported poorer adherence on most process measures (with the notable exception of median time to ECG), compared with hospitals with lower IQI 15 rates. As expected, teaching hospitals, high-volume hospitals for AMI care, and high-volume hospitals for PCI, had much lower risk-adjusted inpatient

mortality, according to IQI 15, than comparator facilities. Hospitals with higher risk-adjusted inpatient mortality, according to IQI 15, also had higher risk-standardized mortality for Medicare FFS patients with AMI, and higher risk-adjusted, all-payer inpatient mortality for CABG, PCI, and HF, compared with hospitals with lower IQI 15 rates. In other words, IQI 15 has the expected relationship with other IQIs at the hospital level.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. 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*)

Using the 2012 data from 36 states, we examined the percent of potential denominator cases excluded by each criterion. The percent of potential cases excluded by each criterion are reported.

2b3.2. 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*)

see Table 3

Table 3. Number and Percent of Discharges Excluded, by Denominator Exclusion Criteria, Acute Myocardial Infarction Mortality Rate (IQI 15)[1]

IQI15	Denominator			Potential Numerator[2]		
Exclusion Name	Exclusion Count	After Exclusions	% Change	Exclusion Count	After Exclusions	% Change
No Exclusions Applied	---	492,529	--	--	25,287	--
Missing DISP	66	492,563	0.0%	0	25,287	0.0%
Transfer to another Acute-care Hospital	43,450	449,079	8.8%	0	25,287	0.0%

[1]This indicator does not have numerator exclusion criteria.
[2]Potential numerator cases are those that would have qualified for the numerator if not for a particular denominator exclusion criterion.
Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 5.0)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The exclusion of patients with missing disposition is necessary because the outcome of the hospital stay (i.e., death versus survival) is unknown for these patients. The prevalence of missingness (approximately 0.01%) is

too low to justify multiple imputation, and stakeholders are unlikely to accept imputation of such an important outcome.

The exclusion of patients transferred to other acute care hospitals is necessary because most users of IQI 15 do not have the ability to link sequential hospitalizations on the same patient to ascertain the ultimate outcome of the inpatient episode of care. To avoid double counting the same AMI, it is necessary to exclude either the initial hospitalization or the subsequent (i.e., second) hospitalization for patients who are transferred from one hospital to another. Given that the outcome of the second hospitalization is known, and given that the second hospital is usually the hospital that provided the largest portion of the patient's care, it is most logical to exclude the initial hospitalization and then to adjust statistically for any difference in risk associated with transfers at the second hospital. (Although there is hypothetical concern that academic medical centers and other hospitals that receive AMI transfers might be disadvantaged by this approach, the empirical data summarized in 2b.2.3 above and Table 4 below do not support this concern. In fact, the Emergency Medical Treatment and Active Labor Act [EMTALA] generally requires that any patient with an "emergency medical condition," such as AMI, be stabilized prior to transfer, meaning that that "no material deterioration of the patient's condition is likely to result from the transfer or is likely to occur during the transfer." Consistent with this standard, the post-transfer, risk-adjusted outcomes of transferred AMI patients are not worse than the outcomes of AMI patients who are not transferred.)

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with 23 risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Clinical Factors
For the provider level indicators, each module has a standard set of covariates grouped into four categories: demographics, severity of illness, comorbidities and transfer-in status. The standard set is tailored to each indicator to create a parsimonious set of covariates for each indicator. Based on cross tabulations between each covariate and the outcome of interest, only those covariates with at least 30 cases with the outcome of interest are retained. For categories that are mutually exclusive, covariates with fewer than 30 cases are pooled into the next covariate along the risk gradient. For example, age 70 to 74 is combined with age 65 to 69, or risk of mortality subclass 3 is combined with subclass 2. For categories with no risk gradient, covariates are pooled into broader covariates. Covariates that are considered as potential risk adjusters for the Inpatient Quality Indicators include gender and age, Major Diagnostic Categories (MDC), All Payer Refined Diagnosis Related Group (APR-DRG), patient point-of-origin and whether they were transferred from another facility.

The omitted covariate within mutually exclusive categories is the reference group for those categories. Reference categories are usually 1) the most common and/or 2) the least risk.

The choice of omitted reference category does affect how one might use the model coefficients or odds ratios in an English language sentence, but it does not affect predicted probabilities or model performance.

Once the preliminary multivariable model is specified, it is estimated on the adult analytic data, as appropriate. Only those covariates that are statistically significant ($p<0.05$) are retained. For covariates that are not statistically significant in categories that are mutually exclusive, the pooling process described above is repeated until a complete, parsimonious model is specified.

Additional details are available in the AHRQ Quality Indicator Empirical Methods document, included in the supplemental file.

Sociodemographic Factors

The relevant literature on disparities in cardiovascular care was recently summarized by Lewey and Choudhry (Lewey J, Choudhry NK. The current state of ethnic and racial disparities in cardiovascular care: lessons from the past and opportunities for the future. Curr Cardiol Rep 2014;16(10):530). These authors make several key points:

1. "black and Hispanic patients presenting with an acute myocardial infarction (AMI) have significantly longer door-to-balloon times compared to white patients, a delay that contributes significantly to clinical outcomes [33, 34]. Differences in utilization are not fully explained by patient preference [35] or overuse among whites [36]."
2. "Several studies have demonstrated that black and Hispanic patients are less likely to receive evidence-based medications and counseling during an admission or at discharge for CHF or AMI. However, many of the differences are attenuated when confounders, such as socioeconomic status, comorbid conditions, and especially site of care, are taken into consideration [4, 14, 33, 43, 44]. Uptake of newer therapies and recommendations may occur more slowly among minority populations. The CRUSADE registry demonstrated that black patients were significantly less likely than whites to receive newer therapies for acute coronary syndrome, such as clopidogrel and glycoprotein IIb/IIIa inhibitors [29].
3. Despite living closer to higher quality hospitals than whites, black patients presenting with AMI or undergoing cardiac surgery are more frequently admitted to lower-quality hospitals [53] and experience higher mortality rates [54–57].
4. Cohen et al. evaluated more than 140,000 patients with AMI treated at one of the 443 participating (Get with The Guidelines, GTWG) hospitals between 2002 and 2007 [33]. During the first year of the study, black patients received lower quality of care compared to whites. But, by the end of the study, the small difference in quality had been eliminated such that all patients were receiving similar quality of care. Despite these improvements, some disparities do persist. A subsequent analysis of the GWTG-CAD cohort found that although time to revascularization in patients presenting with AMI decreased over time, black and Hispanic men continued to be at greater risk for a door-to-balloon time of greater than 90 min [89]."

These findings are supported by a more recent study by Agarwal et al. (Agarwal S, Garg A, Parashar A, et al. Outcomes and resource utilization in ST-elevation myocardial infarction in the United States: Evidence for socioeconomic disparities. J Am Heart Assoc 2014;3:e001057), not included in Lewey and Choudhry's review, that used the same HCUP data on which IQI 15 was developed and tested (specifically, the 2003-2011 Nationwide Inpatient Sample). These authors divided STEMI patients into quartiles based on the median household income of each patient's residential zip code. There was significantly higher inpatient risk-adjusted mortality among the lowest SES quartile as compared to the highest quartile (OR [95% CI]: 1.11 [1.06-1.17]). Similarly, there was a highly significant trend indicating progressively reduced timely reperfusion (i.e., day 0) among patients from lower quartiles (OR [95% CI]: 0.80 [0.74-0.88] comparing the lowest quartile to the

highest). Paradoxically, there was also lower utilization of circulatory support devices among patients from lower as compared to higher zip code quartiles (OR [95% CI]: 0.85 [0.75-0.97] comparing the lowest quartile to the highest).

Based on the evidence thoroughly summarized by Lewey and Choudhry, and recently extended by Agarwal et al., it appears likely that any relationship between socioeconomic factors (e.g., race, ethnicity, income) and inpatient AMI mortality is at least partially mediated by the quality of care provided. Although we cannot exclude the possibility that part of this relationship is independent of quality of care, or is mediated by pre-hospital care (which may not fall within the proper realm of hospital accountability), we have no mechanism by which to separate this component of the socioeconomic effect. Accordingly, consistent with the guidance provided by NQF in the SDS Trial Period FAQs, AHRQ believes that it would be inappropriate to include other SDS variables in the risk-adjustment approach for IQI 15, which is an in-hospital outcome measure.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

The process to select risk factors is described in the AHRQ QI Empirical Methods report. The results of the analyses are provided in the PSI Parameter Estimates document. Both documents are available to reviewers in the supporting materials. The results of the analyses are provided in the tables below as well as on the submitted excel spreadsheet.

There are several steps involved in estimating the QI risk-adjustment models.
1. Construct candidate covariates
2. Select model covariates
3. Estimate the models
4. Evaluate the models

Covariates are coded for each discharge record based on the data elements, data values, and logic described in the technical specifications and the appendices of the risk-adjustment coefficient tables. For a given covariate, if the discharge meets the technical specification for that covariate a value of "1" is assigned to the discharge level covariate data element. Otherwise a value of "0" is assigned to the discharge level covariate data element.

Table 4. Risk Adjustment Coefficients for IQI 15 Acute Myocardial Infarction (AMI) Mortality

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
Intercept		1	-7.1165	0.1201	3512.277	<.0001
Age	18 to 39	1	-0.4068	0.0899	20.4685	<.0001
Age	40 to 44	1	-0.2189	0.0706	9.6089	0.0019
Age	45 to 49	1	-0.2427	0.0549	19.561	<.0001
Age	50 to 54	1	-0.2501	0.0439	32.4341	<.0001
Age	55 to 59	1	-0.1228	0.0383	10.2944	0.0013
Age	65 to 79	1	0.0449	0.0277	2.6264	0.1051
Age	80 to 84	1	0.1283	0.0313	16.7528	<.0001
Age	85+	1	0.3641	0.0294	153.6422	<.0001
APR-DRG	161-(1,2) CARDIAC DEFIBRILLATOR & HEART ASSIST IMPLANT, Risk of mortality (ROM) 1-2	1	2.6906	0.3981	45.6777	<.0001
APR-DRG	161-(3,4) CARDIAC DEFIBRILLATOR & HEART ASSIST IMPLANT, Risk of mortality (ROM) 3-4	1	5.4594	0.1269	1849.559	<.0001

APR-DRG	162-(1,2) CARDIAC VALVE PROCEDURES W CARDIAC CATHETERIZATION, ROM 1 and 2	1	1.7417	0.7185	5.8753	0.0154
APR-DRG	162-3 CARDIAC VALVE PROCEDURES W CARDIAC CATHETERIZATION, ROM 3	1	3.4881	0.2508	193.3492	<.0001
APR-DRG	162-4 CARDIAC VALVE PROCEDURES W CARDIAC CATHETERIZATION, ROM 4	1	5.506	0.1488	1369.386	<.0001
APR-DRG	165-(1,2) CORONARY BYPASS W CARDIAC CATH OR PERCUTANEOUS CARDIAC PROC, ROM 1 and 2	1	1.1177	0.2127	27.616	<.0001
APR-DRG	165-3 CORONARY BYPASS W CARDIAC CATH OR PERCUTANEOUS CARDIAC PROC, ROM 3	1	2.8484	0.1453	384.1976	<.0001
APR-DRG	165-4 CORONARY BYPASS W CARDIAC CATH OR PERCUTANEOUS CARDIAC PROC, ROM 4	1	5.1061	0.1252	1663.658	<.0001
APR-DRG	173-(1-4) OTHER VASCULAR PROCEDURES, ROM 1-4	1	5.1365	0.1574	1065.584	<.0001
APR-DRG	174-2 PERCUTANEOUS CARDIOVASCULAR PROCEDURES W AMI, ROM 2	1	1.622	0.132	150.9876	<.0001
APR-DRG	174-3 PERCUTANEOUS CARDIOVASCULAR PROCEDURES W AMI, ROM 3	1	3.4453	0.124	771.9065	<.0001
APR-DRG	174-4 PERCUTANEOUS CARDIOVASCULAR PROCEDURES W AMI, ROM 4	1	5.9445	0.1192	2488.872	<.0001
APR-DRG	190-1 ACUTE MYOCARDIAL INFARCTION, ROM 1	1	1.6212	0.1564	107.4042	<.0001
APR-DRG	190-2 ACUTE MYOCARDIAL INFARCTION, ROM 2	1	2.9795	0.1249	569.3856	<.0001
APR-DRG	190-3 ACUTE MYOCARDIAL INFARCTION, ROM 3	1	4.316	0.1193	1308.086	<.0001
APR-DRG	190-4 ACUTE MYOCARDIAL INFARCTION, ROM 4	1	6.4592	0.119	2948.145	<.0001
MDC	5 CIRCULATORY SYSTEM, DISEASES & DISORDERS	1	4.768	0.1219	1529.822	<.0001
TRNSFER	TRANSFER STATUS	1	-0.0556	0.0203	7.5194	0.0061

c-statistic=0.887

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Not applicable (see above)

2b4.5. 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*)

This analysis evaluates how strongly the risk adjustment model is associated with the event of interest (i.e. hospitalization for dehydration). The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the c-statistic. The c-statistic is computed by assigning each observation a predicted probability of the outcome from the risk-adjustment model based on the value of the observations covariates from the risk-adjustment model. Two copies of the dataset are sorted, first from highest to lowest predicted probability and second from lowest to highest predicted probability. This creates a set of

pairs of observations. Pairs that consist of one event and one non-event (discordant pairs) are kept and concordant pairs are discarded. The c-statistic is a measure of the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the non-event. C-statistics above 0.70 and below 0.80 have moderate discrimination. Above 0.80 the discrimination is high. We did not employ common "goodness of fit" tests because these tests tend to not be informative with large samples.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

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

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Table 5. Risk adjustment Model Discrimination and Calibration IQI 15 Acute Myocardial Infarction (AMI) Mortality Rate

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate	Observed Rate
1	44,155	0.000641	0.000249
2	44,156	0.000808	0.000815
3	44,156	0.002442	0.002084
4	44,155	0.003837	0.003216
5	44,156	0.006647	0.006975
6	44,156	0.016914	0.017801
7	44,155	0.041646	0.041218
8	44,156	0.069691	0.066356
9	44,157	0.102888	0.105668
10	44,155	0.318175	0.319307
C-Statistic	0.8867		

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 5.0)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

See Table 5 in 2b4.6

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. 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)

A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles. This indicator is well calibrated and has good discrimination, as the observed to predicted values across the deciles range between 0.85 – 1.04 for all deciles except the lowest decile. For hospitals with very low predicted rates, the relative differences between observed and predicted values is greater, primarily due to the very small observed rates.

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

This analysis assesses the probability that a hospital is higher or lower than a benchmark or threshold, given hospital size. It reflects whether the indicator can discriminate the best performing hospitals from the lower performing hospitals.

For this analysis, "benchmark" refers to the smoothed indicator rate based on the 20th percentile of the reference population (i.e., 20% of hospitals have a lower mortality rate or better performance). "Threshold" refers to the indicator rate based on the 80th percentile (i.e., 80% have lower mortality or better performance).

The analysis is reported by size decile, based on the denominator cases, demonstrating performance across hospitals of various sizes. Each hospital is assumed to have an underlying distribution of smoothed rates that follows a Gamma distribution. The parameters of a Gamma distribution are shape and scale. For each hospital the shape is calculated as ((smoothed rate)^2/ smoothed rate variance), and the scale is calculated as (smoothed rate variance / smoothed rate). The smoothed rate variance (aka posterior variance) is calculated as the signal variance – (reliability weight * signal variance). The reliability weight is calculated as (signal variance / (signal variance + noise variance)). Hospitals are ranked by size and grouped into 10 equal categories of size (deciles). The Benchmark and Threshold are compared to the Gamma distribution of the smoothed rates for each hospital to determine if the hospital rate is better or worse than the Benchmark and Threshold rates with 95% probability. This provides a 95% confidence interval for the Benchmark and Threshold rate.

Table 6 reports the proportion of hospitals above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of hospitals not classified as either better or worse have rates that fall within the 95% confidence interval.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 6. Performance Categories by Hospital Size Decile IQI 15 Acute Myocardial Infarction (AMI) Mortality Rate

	Benchmark	Threshold

Size Decile	Number of Hospitals	Average Number of Denominator Discharges Per Hospital	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1	266	4.2	0.0000	0.5075	0.4925	0.1015	0.1278	0.4060
2	266	8.2	0.0075	0.6090	0.3835	0.3496	0.2669	0.2669
3	267	14.9	0.0936	0.7116	0.1948	0.3933	0.2884	0.4120
4	266	27.7	0.1541	0.6729	0.1729	0.5902	0.2180	0.2368
5	267	51.5	0.1423	0.7228	0.1348	0.6592	0.2472	0.2060
6	266	100.8	0.1617	0.7556	0.0827	0.7669	0.1767	0.1504
7	267	166.2	0.0974	0.8390	0.0637	0.7528	0.1948	0.1835
8	266	249.1	0.0865	0.8910	0.0226	0.8308	0.1391	0.1466
9	267	361.3	0.0562	0.9101	0.0337	0.8876	0.0974	0.0787
10	266	672.2	0.0376	0.9511	0.0113	0.9436	0.0489	0.0451
Overall	2,664	165.6	0.0837	0.7571	0.1592	0.6276	0.1806	0.2132

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 5.0)

2b5.3. 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?*)

This indicator has strong discrimination for most hospitals to identify low performing hospitals; 79% of hospitals can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold). The indicator has strong discrimination, particularly for moderate to large hospitals to identify high performing hospitals; 84% of hospitals can be classified as better or worse than the benchmark.

2b6. 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 SDS 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 SDS 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.***

Not applicable

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not applicable

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

The AHRQ QIs use frequently reported administrative data variables. IQI 15 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all hospital discharge records. The rate of missing data for each variable is available by state and year from the AHRQ HCUP website (http://www.hcup-us.ahrq.gov/cdstats/cdstats_search.jsp).

2b7.2. 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*)

For these variables, rates of missing data are typically less than 1% of the state database. It is unlikely the bias would occur from such a low rate of missing data.

2b7.3. 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*)

Exclusion of cases for missing data is appropriate.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue

burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Because the indicator is based on readily available administrative billing and claims data and U.S. Census data, feasibility is not an issue. This version of the indicator requires POA data for risk-adjustment (although not for specification of the numerator and denominator). Present-on-Admission indicators were added as data elements to the uniform bill form (UB-04) effective October 1, 2007. Hospitals incurred a payment penalty for not including POA status on Medicare records beginning October 1, 2008. Each of the secondary diagnoses in a discharge record can be flagged as "present at the time the order for inpatient admission occurs" or not (see http://www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm). The number of states reporting consistent POA has increased dramatically since 2008.

The AHRQ QI software has been publicly available at no cost since 2001; Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

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)*.

There are no fees. The IQI15 software is freely available from the AHRQ Quality Indicators website (http://www.qualityindicators.ahrq.gov/).

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.

Planned	Current Use (for current use provide URL)
	Public Reporting Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website http://pub.azdhs.gov/hospital-discharge-stats/2012/index.html http://pub.azdhs.gov/hospital-discharge-stats/2011/index.html Arkansas Department of Health, Arkansas Hospital Discharge Health Data Site (MONAHRQ-generated) http://healthdata.ar.gov/Methodology.html http://hcupnet.ahrq.gov/HCUPnet.jsp?Id=C3A7C545514B2C08&Form=MAINSEL&JS=Y&Action=%3E%3ENext%3E%3E&_MAINSEL=AHRQ%20Quality%20Indicators California Office of Statewide Health Planning & Development http://www.oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/iqi-imi_overview.html Commonwealth Fund, Why Not the Best http://www.whynotthebest.org/measures/view/10858#default&measure=10858&hidemeasures=1&unit=county Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website http://ctmonahrq.ct.gov/2012/index.html#/quality-ratings HealthGrades http://www.healthgrades.com/quality http://www.healthgrades.com/quality/2014-patient-safety-methodology http://www.healthgrades.com/ratings-and-awards/data-source-patient-safety Illinois Department of Public Health http://www.healthcarereportcard.illinois.gov/measures/view/10192/101262%7C101240%7C101165 http://www.healthcarereportcard.illinois.gov/ Kentucky Health Care Information Center, MONAHRQ website http://chfs.ky.gov/ohp/healthdata https://prd.chfs.ky.gov/MONAHRQ/2012/MONAHRQ/AboutQualityRatings.html#L Maine Health Data Organization (MHDO), MONAHRQ Website http://gateway.maine.gov/mhdo/monahrq/index.html http://gateway.maine.gov/mhdo/monahrq/Methodology.html Maryland Health Care Commission, MONAHRQ Website https://www.marylandqmdc.org/ Nevada Compare Care, MONAHRQ website http://nevadacomparecare.net/ http://nevadacomparecare.net/Monahrq/index.html#/resources/Definitions http://nevadacomparecare.net/Monahrq/AboutQualityRatings.html http://nevadacomparecare.net/Monahrq/index.html#/ New York State Department of Health http://profiles.health.ny.gov/measures/all_state/16773 Oklahoma State Department of Health, MONAHRQ website http://www.ok.gov/health/pub/wrapper/ok2share.html Pennsylvania Health Care Cost Containment Council http://www.phc4.org/hpr/Results.aspx?Years=20092-20101&CC=AMI_Medical&CID=0&Facilities Utah Department of Health, MONAHRQ website https://health.utah.gov/myhealthcare/monahrq/ https://health.utah.gov/myhealthcare/monahrq/AboutQualityRatings.html#Q Virginia Health Information, MONAHRQ website http://www.vhi.org/monahrq2/qual/PHC/maps/s_All.html http://www.vhi.org/MONAHRQ/default.asp?yr=2013 https://health.utah.gov/myhealthcare/monahrq/AboutQualityRatings.html

	Washington State, MONAHRQ website http://www.wamonahrq.net/ http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/Definitions Quality Improvement (Internal to the specific organization) University HealthSystem Consortium https://www.uhc.edu https://www.uhc.edu/22982 https://www.uhc.edu/performance-intelligence

4a.1. For each CURRENT use, checked above, provide:

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

Public Reporting:
Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Arizona
http://pub.azdhs.gov/hospital-discharge-stats/2012/index.html
http://pub.azdhs.gov/hospital-discharge-stats/2011/index.html

Arkansas Department of Health, Arkansas Hospital Discharge Health Data Site (MONAHRQ-generated)
County-level hospital admission rate data from most hospitals in Arkansas
http://healthdata.ar.gov/Methodology.html
http://hcupnet.ahrq.gov/HCUPnet.jsp?Id=C3A7C545514B2C08&Form=MAINSEL&JS=Y&Action=%3E%3ENext%3E%3E&_MAINSEL=AHRQ%20Quality%20Indicators

California Office of Statewide Health Planning & Development
Hospital quality ratings from all hospitals in California
http://www.oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/iqi-imi_overview.html

Commonwealth Fund, Why Not the Best
Compares hospital performance at the county, HRR, state, and national levels
http://www.whynotthebest.org/measures/view/10858#default&measure=10858&hidemeasures=1&unit=county

Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Arizona
http://ctmonahrq.ct.gov/2012/index.html#/quality-ratings

HealthGrades
Healthgrades measures 40 million patient records from 4,500 hospitals nationwide for the most recent three-year period. Consumer-targeted hospital and provider ratings.
http://www.healthgrades.com/quality
http://www.healthgrades.com/quality/2014-patient-safety-methodology
http://www.healthgrades.com/ratings-and-awards/data-source-patient-safety

Illinois Department of Public Health
Illinois Hospital Report Card
http://www.healthcarereportcard.illinois.gov/measures/view/10192/101262%7C101240%7C101165
http://www.healthcarereportcard.illinois.gov/

Kentucky Health Care Information Center, MONAHRQ website
Quality reporting on hospitals across the state of Kentucky (n = 132 in 2012).
http://chfs.ky.gov/ohp/healthdata
https://prd.chfs.ky.gov/MONAHRQ/2012/MONAHRQ/AboutQualityRatings.html#L

Maine Health Data Organization (MHDO), MONAHRQ Website
Hospital quality ratings from all hospitals in Maine
http://gateway.maine.gov/mhdo/monahrq/index.html
http://gateway.maine.gov/mhdo/monahrq/Methodology.html

Maryland Health Care Commission, MONAHRQ Website
Compares quality ratings on hospitals across Maryland
https://www.marylandqmdc.org/

Nevada Compare Care, MONAHRQ website
Hospital quality ratings from most hospitals in Nevada: Quality reporting on hospitals across the state of Nevada Under NV Regulation R151-8 this transparency website presents hospital quality and utilization information
http://nevadacomparecare.net/
http://nevadacomparecare.net/Monahrq/index.html#/resources/Definitions
http://nevadacomparecare.net/Monahrq/AboutQualityRatings.html
http://nevadacomparecare.net/Monahrq/index.html#/

New York State Department of Health
The Provider Profiles website displays valuable information about healthcare providers in New York State.
http://profiles.health.ny.gov/measures/all_state/16773

Oklahoma State Department of Health, MONAHRQ website
Includes vital statistics, hospital and ASC discharges, health surveys, and health registries for hospitals in the state of Oklahoma
http://www.ok.gov/health/pub/wrapper/ok2share.html

Pennsylvania Health Care Cost Containment Council
Collects 4.5 million inpatient hospital discharge and ambulatory/outpatient procedure records each year from hospitals and freestanding ambulatory surgery centers in Pennsylvania.
http://www.phc4.org/hpr/Results.aspx?Years=20092-20101&CC=AMI_Medical&CID=0&Facilities

Utah Department of Health, MONAHRQ website
Hospital quality ratings from all hospitals in Utah
https://health.utah.gov/myhealthcare/monahrq/
https://health.utah.gov/myhealthcare/monahrq/AboutQualityRatings.html#Q

Virginia Health Information, MONAHRQ website
Compares quality ratings on hospitals across Virginia.
http://www.vhi.org/monahrq2/qual/PHC/maps/s_All.html
http://www.vhi.org/MONAHRQ/default.asp?yr=2013
https://health.utah.gov/myhealthcare/monahrq/AboutQualityRatings.html

Washington State, MONAHRQ website
Information system of inpatient care utilization, quality, and potentially avoidable stays in Washington State's community hospitals.
http://www.wamonahrq.net/
http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/Definitions

Quality Improvement:

University HealthSystem Consortium
Internal quality improvement efforts including Postoperative Respiratory Failure 2008 Benchmarking Project and documentation and evaluation of AHRQ PSIs for quality improvement by its members.
Internal quality improvement efforts including Postoperative Respiratory Failure 2008 Benchmarking Project.
https://www.uhc.edu
https://www.uhc.edu/22982
https://www.uhc.edu/performance-intelligence

4a.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?*)
Not applicable

4a.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.*)
Not applicable

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

See Table 1 in response to question 1b.2.

The observed rate has decreased over time, from 68.9 cases per 1000 in 2008 to 56.4 cases per 1000 in 2012. The mean hospital-level rate in 2012 was 104.3 with a coefficient of variation of 1.44, and 111.90 in 2011 with a coefficient of variation of 1.38. The coefficient of variation suggests that further improvement may be possible if lower performing hospitals met the performance of higher performing hospitals.

These rates encompass 82 percent of all U.S. community hospital discharges from 36 of the 45 states that participated in the 2012 SID, for a total of about 30 million hospital discharges from community hospitals, which include all non-Federal, short-term, general or other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are public and academic medical centers, specialty hospitals such as obstetrics–gynecology, ear–nose–throat, orthopedic and pediatric institutions. Short-stay rehabilitation and long-term acute care hospitals are excluded from the data used for the reported analyses. The remaining discharges not captured in these data are psychiatric facilities, alcohol and drug dependency facilities and military hospitals. The number and percentage of discharges captured in past years are similar. For IQI 15, 24,890 patients died among 441,557 AMI patients at 2,978 hospitals in 2012.

4b.2. **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.**
Not applicable

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
None identified or reported. In theory, the use of this metric could encourage premature transfer of dying patients to lower levels of care, such as skilled nursing or intermediate care facilities, but the widespread use of 30-day risk-standardized mortality measures for Medicare beneficiaries(which are immune to this unintended consequence) makes such behavior unlikely.

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)**
0230 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older
2473 : Hospital 30-Day Risk-Standardized Acute Myocardial Infarction (AMI) Mortality eMeasure

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

5a. Harmonization
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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
The indicators referenced above include 30-day mortality 1) for patients age 18 years and older 2) specified as an e-measure and 3) for patients age 65 and older. Inpatient mortality and 30-day mortality are different concepts, although capturing the same ultimate outcome. Harmonization is not appropriate.

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.)
IQI 15 and the Centers for Medicare & Medicaid Services' NQF-endorsed measures concerning AMI mortality (0230 and 2473) use the same ICD-9-CM codes to identify AMI, but they differ in two important respects: (1) whereas the CMS measures concern only Medicare fee-for-service and VA beneficiaries 65 years or older, IQI 15 measures mortality among hospitalizations of patients 18 years or older at non-federal acute care hospitals for all payers; and (2) while the CMS measures evaluate 30-day mortality, IQI 15—because it is based only on UB-04 data elements—is limited to inpatient mortality. The latter difference is a potential disadvantage in that the time at risk is not uniform for all patients and 30-day mortality is typically greater than inpatient mortality, but the former difference is an advantage because IQI 15 encompasses a greater proportion of the entire population at risk. We therefore believe that #0730 complements #0230 by offering an alternative specification for users who are interested in patients of all ages and all payers, just as #2473 offers an alternative e-measure specification for those with electronic health data.

Appendix

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

Attachment **Attachment:** IQI15_Supplemental_Files.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Agency for Healthcare Research and Quality
Co.2 **Point of Contact:** Mamatha, Pancholi, Mamatha.Pancholi@ahrq.hhs.gov, 301-427-1470-
Co.3 **Measure Developer if different from Measure Steward:** Agency for Healthcare Research and Quality
Co.4 **Point of Contact:** Carol, Stocks, Carol.Stocks@ahrq.hhs.gov, 301-427-1422-

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.
N/A

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:** 03, 2011
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?** 10, 2015

Ad.6 **Copyright statement:** The AHRQ QI software is publicly available. We have no copyright disclaimers.
Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 0965
Measure Title: Discharge Medications (ACE/ARB and beta blockers) in Eligible ICD Implant Patients
Measure Steward: American College of Cardiology
Brief Description of Measure: Proportion of patients undergoing ICD implant who received prescriptions for all medications (ACE/ARB and beta blockers) for which they are eligible for at discharge.
Developer Rationale: This measure is intended to assess the extent to which eligible patients receive evidence-based medications that are indicated at hospital discharge following ICD implantation.
Composite performance measures have a variety of uses.
Data reduction. A large and growing array of individual indicators makes it possible for users to become overloaded with data. A composite measure reduces the information burden by distilling the available indicators into a simple summary.
Scope expansion. The information in a composite measure is highly condensed, making it feasible to track a broader range of metrics than would be possible otherwise. Composite measures have been described as a tool for making provider assessments more comprehensive
Provider performance valuation. Performance indicators are used for various decisions about providers, including the allocation of pay-for-performance incentives, designation of preferred provider status, and assignment of letter grades and star rating categories. If a decision is to be based on multiple indicators instead of a single indicator, a method of translating several variables into a single decision is needed. Composite measures serve this function by assigning providers to 1 position on a scale of better-to-worse performance.
Given all these uses, NCDR believes that while we will continue to report these measures at the individual level there is a distinctive value of having a composite measure endorsed at NQF.

Numerator Statement: Patients who receive ACE/ARB and Beta blockers for which they are eligible.
1. ACE/ARB prescribed at discharge (if eligible for ACE/ARB as described in denominator)
AND
2. Beta blockers prescribed at discharge (if eligible for beta blockers as described in denominator)

Denominator Statement: All patients with an ICD implant surviving hospitalization who are eligible to receive any one of the two medication classes:
1) Eligiblility for ACE/ARB: Patients who have an ejection fraction (EF) of <40% AND do not have a documented contraindication to ACE/ARB documented
OR
2) Eligibility for beta blockers: Patients who do not have a documented contraindication to beta blocker therapy and have either:
a. EF of <40% OR
b. a previous myocardial infarction (MI)

Denominator Exclusions: Discharge status of expired; not eligible for either ACE/ARB or beta blockers

Measure Type: Composite

Data Source: Electronic Clinical Data : Registry

Level of Analysis: Facility

Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

IF this measure is included in a composite, NQF Composite#/title: n/a – 0965 is the composite measure.

Composite Measure Construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)

Component Measures (if endorsed or submitted for endorsement): n/a

The *non-endorsed* component measures for this composite measure include:

1. ACE/ARB prescribed at discharge (if eligible for ACE/ARB as described in denominator)
2. Beta blockers prescribed at discharge (if eligible for beta blockers as described in denominator)

Is this a MAINTENANCE measure submission? ☒ **Yes** ☐ **No, this is a NEW measure submission.**

For MAINTENANCE, state the Original Endorsement Date: 1/17/12 **Most Recent Endorsement Date:** 1/17/12

Previous Measure Evaluation - Public & Member Comments, Developer Responses & Steering Committee Recommendations from (Cardiology Project 2010):

Public and Member Comments:

- This measure is too specific to be generalized to the population.

Steering Committee:

- ICD patients are an important population that has a special clinical registry to track the performance. This all-or-none composite measure was specifically developed at the request of the Steering Committee to increase the number of composite measures.

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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.

- This composite measure has two process measure components. Evidence for each component measure should include a systematic review of the body of evidence that describes the quantity, quality and consistency (QQC) of the evidence that relates the process of care to desired health outcomes. Evidence for this measure should link receipt of each medication to patient outcomes. (algorithm Box 3)
- This composite measure has 2 component measures that assess if all patients with an ICD implant surviving hospitalization who are eligible to receive any one of the two medication classes, beta-blockade and ACE/ARBs. Because the beta-blocker component may be applied to two separate patient populations (patients with previous MI and patients with LVSD), the developer has provided evidence supporting the use of beta blockers in each of these populations separately. The developer provides diagrams demonstrating how receiving beta-blockers for a previous MI, LVSD and ACEI/ARBs for LVSD are linked to patient outcomes.

Beta-blocker for previous MI

- The developer provides 4 guidelines with 6 guideline statements that recommend beta-blocker therapy for patients with HF or prior MI and the Quantity/Quality/Consistency for the 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline (Level of Evidence A or B).
- One prospective cohort study and one meta-analysis were published after the publication of the 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline. The analysis concluded that the use of beta-blockers in patients with stable CAD was associated with a lower risk of cardiovascular mortality.

Beta-blocker for LVSD

- The developer provides 2 guidelines with 4 guideline statements that recommend beta-blocker therapy for patients with LVSD, with or without prior MI and the Quantity/Quality/Consistency for the 2013 ACCF/AHA Guideline for the Management of Heart Failure (Level of Evidence A, B, or C).
- One RCT, one prospective cohort study, and two meta-analyses were published after the publication of the 2013 ACCF/AHA Guideline for the Management of Heart Failure. The analysis concluded that the use of beta-blockers in patients with stable CAD was associated with a lower risk of cardiovascular mortality.

ACE/ARBs for LVSD

- The developer provides 2 guidelines with 6 guideline statements that recommend that recommend ACE/ARBs for patients with LVSD, with or without prior MI and the Quantity/Quality/Consistency for the 2013 ACCF/AHA Guideline for the Management of Heart Failure (Level of Evidence A or B).
- One meta-analysis was published after the publication of the 2013 ACCF/AHA Guideline for the Management of Heart Failure.

Questions for the Committee:

- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- In 2011-2012 a total of 243,186 patients at 1552 hospitals were analyzed and 195,563 patients at 1606 hospitals in 2013-14. Data from 2011-12 indicated a mean of 74% and 50th percentile results at 76%. Data from 2013-14 indicated a mean of 78% and 50th percentile results at 79%.
- The developers currently collect data on race and insurance type but do not provide statistical analysis and results of these data elements. Instead, the developer provides the # and % of the following hospital characteristics by age, sex, race and insurance type.
- The developers do not provide a summary from the literature that addresses disparities in care on the specific focus of this measure.
- The developers provide the following rationale for the composite:
  - This measure is intended to assess the extent to which eligible patients receive evidence-based medications that are indicated at hospital discharge following ICD implantation.
  - Each of the components of this measure address appropriate medication prescribing at discharge for ICD patients. Combining the individual process measures into a single composite provides patients, physicians, and hospitals with a perspective of the overall quality of medical therapy provided to patients undergoing ICD implantation.

- The composite measure is an "all-or-none measure" meaning all essential care processes are received by each patient.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - I think that there is not significant evidence to support this measure. There is very good evidence that patients with AMI and depressed LV function or CAD and depressed LV fxn benefit from beta blockers, but there is not significant evidence in a purely ICD population of patients.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- I think that there is not significant evidence to support this measure. There is very good evidence that patients with AMI and depressed LV function or CAD and depressed LV fxn benefit from beta blockers, but there is not significant evidence in a purely ICD population of patients.

1b. ***Committee's Comments on Performance Gap:***

- There is a performance gap

1c. ***Committee's Comments on Composite Performance Measure:***

- I could go either way on this one. It is not clear to me there is a significant advantage of a composite measure.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure's data source is the National Cardiovascular Data Registry (NCDR) ICD Registry with the data dictionary and collection tool provided. The data elements and the calculation algorithm are described. The developer does not provide ICD-9, ICD-10 codes and the specific beta-blockers and ACE/ARBs are not specified - this was brought up in CV2 with 0964. The developer explained that the measure aligns with the guidelines which are at a drug class-level. There is no specific recommendation in the guidelines for one BB/ACE/ARB over another.
- Denominator exclusions include "discharge status of expired patients" and "not eligible for either ACE/ARB or beta-blockers". "Not eligible" includes patients in a clinical trial and with a contraindication; types of acceptable contraindications are not provided.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*

○ *Is the logic or calculation algorithm clear?*

○ *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- Empirical testing was carried out using the National Cardiovascular Data Registry for ICD Registry from 1,606 hospitals. The patient characteristics are described.
- Reliability testing of the measure score was performed using a correlation of random split halves. The correlation coefficient of 0.87 is high.
- The developers provide a description of the registry's Data Quality Program that includes onsite audits and Inter-Rater Reliability Assessment conducted to validate the audit. The kappa scores were calculated with a 95% CI; a kappa > .70 is considered acceptable inter-rater reliability. The developers state that "The kappa score for all medication elements demonstrate substantial or almost perfect reliability."

Questions for the Committee:

○ *Is the test sample adequate to generalize for widespread implementation?*

○ *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The specifications include the specific medications described in the evidence for PCI patients.

Question for the Committee:

○ *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- Empiric testing was conducted at the level of the data element and measure score using 93,971 Medicare FFS patients who were at least 65 years of age and underwent ICD implantation in 2010 or 2011.
- The analyses included the association of patient and hospital performance on the composite measure with adverse outcomes, specifically mortality and readmission at 6 months following hospital discharge and the association between hospital-level performance on the measure and the combination of mortality or readmission at 6 months. The developer provides patient-level and hospital level results:
    - A significantly smaller proportion of patients discharged on the appropriate medical therapy died or were readmitted within 6 months of hospital discharge (without meds = 28.37% vs. with meds = 36.28%).
    - Patients treated at hospitals that performed better on the measure had better unadjusted outcomes that those treated at hospitals that performed worse on the measure (correlation coefficient (-0.0998), $p<0.001$).
- Face validity was described as "Content validity of this process was achieved by the specialized expertise" of various ACC committee members involved in the development or approval of the measure.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The only exclusions for this measure are noted under S.10. (Discharge status of expired; not eligible for either ACE/ARB or beta blockers). These exclusions are relatively rare and firmly supported by the clinical rationale.
- The developer provides data on the frequency of exclusions.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- The measure is not risk-adjusted.
- The developer states that they do not currently collect many of the socio-demographic status (SDS) variables listed as examples in the submission form, but note that they do collect data on race as well as insurance type.
- However, the developer did not consider either clinical or SDS adjustment for this measure.

Questions for the Committee:

- *Do you agree with the developer that risk-adjustment for clinical or SDS factors is not necessary for this measure?*

2b5. Meaningful difference:

- Detailed performance results are presented for many subpoulations. Results vary among different groups.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- The developer reports that missing data defaults to performance not met. This measure assumes that missing documentation on the process results in a failure of meeting evidence based therapy.

2d.Composite measure: construction

- The developer reports that the empirical analysis demonstrating the individual component measures fit the overall quality construct is currently being researched. This research is expected to be published in the medical literature.

Questions for the Committee:

- *Are the quality construct and rationale for the composite explicitly stated and logical?*
- *Is the method for aggregation and weighting of the components explicitly stated and logical?*

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- This would be enhanced if they used ICD-9 or ICD10 codes but they have been capturing everything in NCDR so I think it meets the bar.

2a2.: ***Committee's Comments on Reliability-Testing:***

- The developers state that there is substantial or almost perfect reliability

2b1.: ***Committee's Comments on Validity-Specifications***:

- This is based upon face validity and i agree it meets that criteria

2b2.: ***Committee's Comments on Validity-Testing:***

- I have no concerns regarding validity.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- My biggest concern here is the failure to adjust for SDS. Patients may or may not be able to afford these classes of drugs and without adjustment for factors related to SDS at an institution, there would be a reporting bias.

2d.: ***Committee's Comments on Composite Performance Measure:***

- They say "currently being researched"

Criterion 3. Feasibility

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

- The developer reports that the data are available via several methods: electronic transfer to the registry from the procedure/care setting; web-based tool for manual data entry or from an EHR.
- The developer states that centers already have to participate in this specific registry for reimbursement purposes therefore there is no additional cost and there is no charge for a standard export package.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- Seems fine and easy enough

Criterion 4: Usability and Use

4. Usability and 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.

- The measure is not currently reported publically. The individual component measures are used in ACC's NCDR Registry.
- Planned use: public reporting, QI with external benchmarking to multiple organizations and internal to the

specific organization.

- Improvement: The developers note that comparing 2011-2012 data with 2013-2014 data – "While the top 10% of performers saw slight performance improvement, the hospitals in the lower percentiles (below the median) improved significantly."
- Unintended consequences: The developer states that inaccuracies may occur due to incorrectly exported data and during the transmission of data from medical record to paper form and/or online data collection tool and some sites may over-code medication exclusions.

***Questions for the Committee**:*

- *Is the measure publicly reported?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- No issues

Criterion 5: Related and Competing Measures

- 0066 : Coronary Artery Disease (CAD): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy - Diabetes or Left Ventricular Systolic Dysfunction (LVEF < 40%)
- 0070 : Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF < 40%)
- 0071 : Persistence of Beta-Blocker Treatment After a Heart Attack
- 0081 : Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)
- 0083 : Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)
- 0117 : Beta Blockade at Discharge
- 0236 : Coronary Artery Bypass Graft (CABG): Preoperative Beta-Blocker in Patients with Isolated CABG Surgery
- 0594 : Post MI: ACE inhibitor or ARB therapy
- 0696 : The STS CABG Composite Score (Composite Measure)
- The developer did not harmonize measure specifications and report that the listed measures are not in direct competition with 0965.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1528

Measure Title: Beta Blocker at Discharge for ICD Implant Patients with a Previous MI

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: 965 Patients with an ICD implant who receive ACE-I/ARB and beta blocker therapy at discharge

Date of Submission: Click here to enter a date

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Beta-blocker therapy for patients with a prior MI receiving an ICD

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

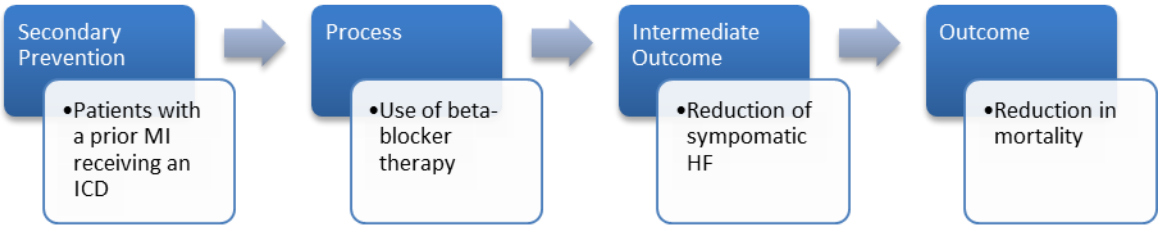
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



Beta-blockers reduce morbidity, mortality, and hospitalizations for patients who had a prior myocardial infarction (MI).

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

_____

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Amsterdam EA, Wenger NK, Brindis RG, Casey DE Jr, Ganiats TG, Holmes DR Jr, Jaffe AS, Jneid H, Kelly RF, Kontos MC, Levine GN, Liebson PR, Mukherjee D, Peterson ED, Sabatine MS, Smalling RW, Zieman SJ. 2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes: a report of the American College of Cardiology/ American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014;64:e139–228.

http://content.onlinejacc.org/article.aspx?articleid=1910086

O'Gara PT, Kushner FG, Ascheim DD, Casey DE Jr, Chung MK, de Lemos JA, Ettinger SM, Fang JC, Fesmire FM, Franklin BA, Granger CB, Krumholz HM, Linderbaum JA, Morrow DA, Newby LK, Ornato JP, Ou N, Radford MJ, Tamis-Holland JE, Tommaso CL, Tracy CM, Woo YJ, Zhao DX. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;61:e78–140, doi:10.1016/j.jacc.2012.11.019.

http://content.onlinejacc.org/article.aspx?articleid=1486115

Smith SC Jr., Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, Gibbons RJ, Grundy SM, Hiratzka LF, Jones DW, Lloyd-Jones DM, Minissian M, Mosca L, Peterson ED, Sacco RL, Spertus J, Stein JH, Taubert KA. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular

disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation. 2011: published online before print November 3, 2011, 10.1161/CIR.0b013e318235eb4d.

http://content.onlinejacc.org/article.aspx?articleid=1147807

Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, Douglas PS, Foody JM, Gerber TC, Hinderliter AL, King SB III, Kligfield PD, Krumholz HM, Kwong RYK, Lim MJ, Linderbaum JA, Mack MJ, Munger MA, Prager RL, Sabik JF, Shaw LJ, Sikkema JD, Smith CR Jr, Smith SC Jr, Spertus JA, Williams SV. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2012;60:e44–164.

http://content.onlinejacc.org/article.aspx?articleid=1391404

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes (p. e 159)

1. In patients with concomitant NSTE-ACS, stabilized HF, and reduced systolic function, it is recommended to continue beta-blocker therapy with 1 of the 3 drugs proven to reduce mortality in patients with HF: sustained-release metoprolol succinate, carvedilol, or bisoprolol. **Class I: Level of Evidence: C**

2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction (p. e104)

2. Beta blockers should be continued during and after hospitalization for all patients with STEMI and with no contraindications to their use. **Class I: Level of Evidence: B**

AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update (p. e2435)

3. Beta-blocker therapy should be used in all patients with left ventricular systolic dysfunction (ejection fraction <40%) with heart failure or prior myocardial infarction, unless contraindicated. (Use should be limited to carvedilol, metoprolol succinate, or bisoprolol, which have been shown to reduce mortality.) **Class I: Level of Evidence: A**
4. Beta-blocker therapy should be started and continued for 3 years in all patients with normal left ventricular function who have had myocardial infarction or ACS. **Class I: Level of Evidence: B**

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease (p. e96)

5. Beta-blocker therapy should be started and continued for 3 years in all patients with normal LV function after MI or ACS. **Class I: Level of Evidence: B**

6. Beta-blocker therapy should be used in all patients with LV systolic dysfunction (EF <40%) with heart failure or prior MI, unless contra- indicated. (Use should be limited to carvedilol, metoprolol succinate, or bisoprolol, which have been shown to reduce risk of death.). **Class I: Level of Evidence: A**

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

Guideline Statement # (see 1a.4.2 above)	Class of Recommendation/Level of Evidence (for definitions see 1a.4.4 below)
1	Class Ic
2	Class Ib
3	Class Ia
4	Class Ib
5	Class Ib
6	Class Ia

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Class of Recommendation (COR) is an estimate of the size of the treatment effect considering risks versus benefits in addition to evidence and/or agreement that a given treatment or procedure is or is not useful/effective or in some situations may cause harm.

Class I: Procedure/Treatment should be performed/administered

Class IIa: It is reasonable to perform procedure/administer treatment

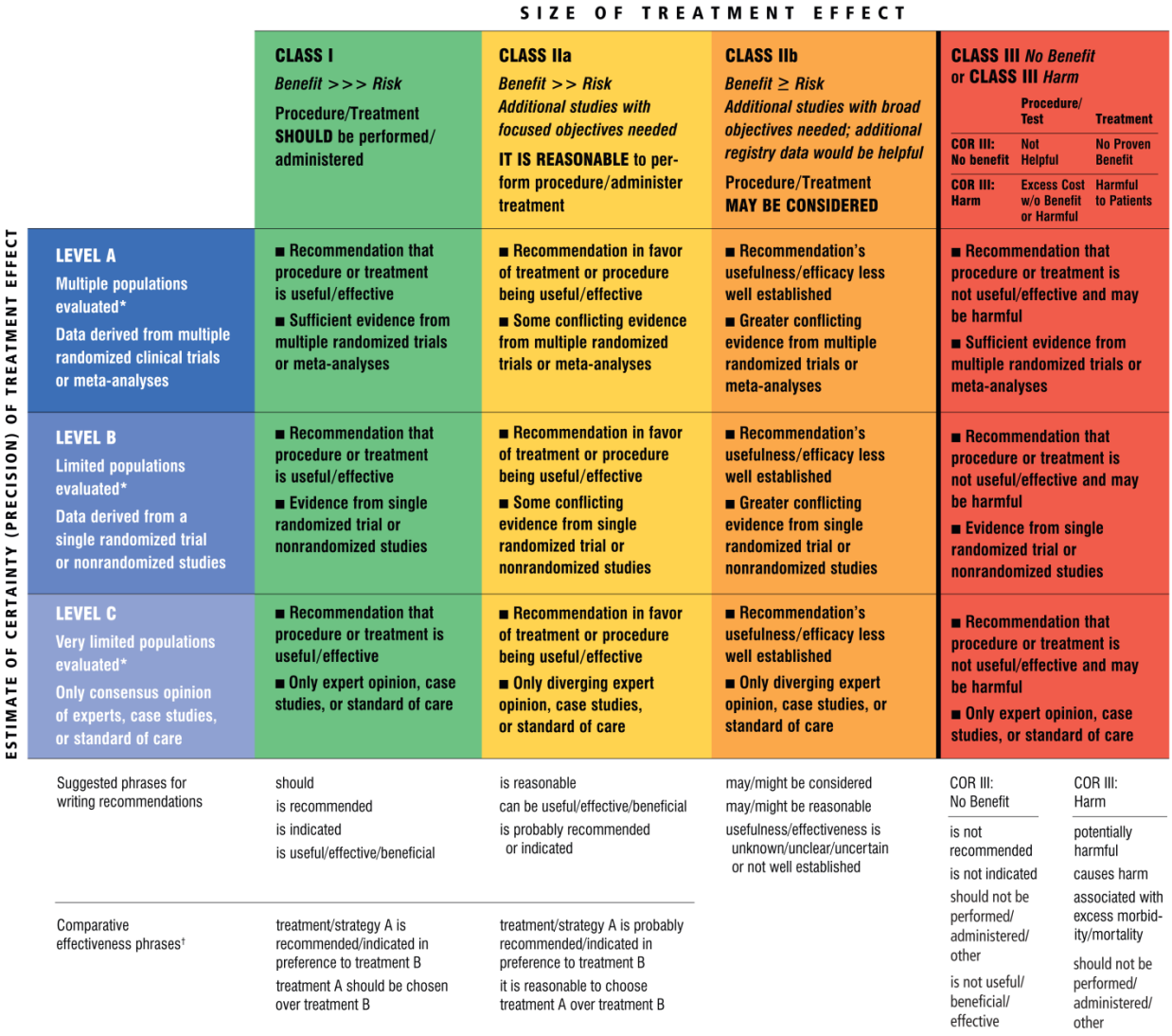
Class IIb: Procedure/Treatment may be considered

Class III: No benefit (Not helpful or No proven benefit)

Class III: Harm (Excess cost w/o benefit or Harmful to patients)

Specific COR definitions are included in Table 1 below.

Table 1. Applying Classification of Recommendation and Level of Evidence



Note: A recommendation with Level of Evidence B or C does not imply that the recommendation is weak. Many important clinical questions addressed in the guidelines do not lend themselves to clinical trials. Although randomized trials are unavailable, there may be a very clear clinical consensus that a particular test or therapy is useful or effective. *Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use. †For comparative effectiveness recommendations (Class I and IIa; Level of Evidence A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

ACCF/AHA Task Force on Practice Guidelines. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines. American College of Cardiology Foundation and American Heart Association, Inc. Cardiosource.com. 2010. Available at:

http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf and

http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease

This guideline covers multiple management issues for the adult patient with stable known or suspected ischemic heart disease (SIHD) including the guideline-directed medical therapy (GDMT) such as beta-blocker therapy.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

An overall grade for the quality of evidence was not assigned. Rather, the quality of a study (or set of studies) supporting a recommendation was graded on an estimate of the certainty or precision of the treatment effect (see 1a.4.3).

Recommendations used to support this measure have a:

- Level of Evidence A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation.

OR

- Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies. References used to determine level of evidence must be provided and cited with the recommendation

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Level of Evidence (LOE) is an estimate of the certainty or precision of the treatment effect.

Level of Evidence A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence C: Consensus opinion of experts, case studies, or standard of care.

Specific LOE definitions are included in Table 1 in 1a.4.4.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: An extensive evidence review was conducted through December 2008 and includes selected other references through December 2011.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease

The body of evidence supporting the recommendations on beta-blocker therapy with patients with a prior MI includes randomized controlled trials and meta-analyses. The number of which is not provided in the guideline.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease

All of the recommendations for this process are rated as Level of Evidence A or B, meaning that the data was derived from one or more RCTs or meta-analyses. Additional information on the overall quality of evidence across the RCTs is not provided.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease (p. e96-97)

Decreases in the rate–BP product, AV nodal conduction, and myocardial contractility from beta blockers reduce myocardial oxygen demand, counteracting beta- receptor activity and contributing to a reduction in angina onset, with improvement in the ischemic threshold during exercise and in symptoms. These agents significantly reduce deaths and recurrent MIs in patients who have suffered a MI and are especially effective when a STEMI is complicated by persistent or recurrent ischemia or tachyarrhythmias early after the onset of infarction. However, no large trials have assessed effects of beta blockers on survival or coronary event rates in patients with SIHD.

Two large long-term follow-up studies investigating the prognostic importance of heart rate showed that all-cause mortality rate progressively increases with higher resting heart rate after adjustment for exercise capacity, age, diabetes mellitus, systolic arterial pressure, BMI, and level of physical activity. Therefore, it is recommended that beta-blocker dosing be adjusted to limit the heart rate to 55 to 60 beats per minute at rest.

In large prospective studies, bisoprolol, carvedilol, and metoprolol, when administered on a background of ACE inhibitors and diuretics with or without digoxin, have been shown to reduce the risk of death and to improve symptoms, clinical status, and quality of life in patients with chronic systolic heart failure. Importantly, these benefits were seen in patients with and without IHD.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease (p. e96-97)

Absolute contraindications to beta blockers are severe bradycardia, preexisting high-degree AV block, sick sinus syndrome (without a pacemaker in place), and refractory heart failure. Relative contraindications include bronchospastic disease or active PAD (beta blockers without vasodilating properties or selective agents at low doses may be used). Because they can mask symptoms of hypoglycemia, beta blockers should be used with caution in patients with insulin-dependent diabetes mellitus. Abrupt beta-blocker withdrawal should be avoided because heightened beta- receptor density and sensitivity can result in a rebound phenomenon associated with an increased risk for AMI and sudden death.

The principle adverse effects of beta blockers are fatigue, exercise intolerance, lethargy, insomnia, nightmares, and impotence.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

One prospective cohort study and one meta-analysis were published after the publication of the 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease.

Note: Text below for description and results is verbatim from the article abstract.

Bangalore S, Makani H, Radford M, Thakur K, Toklu B, Katz SD, DiNicolantonio JJ, Devereaux P, Alexander KP, Wettersley J, Messerli FH. Clinical outcomes with β-blockers for myocardial infarction: a meta-analysis of randomized trials. Am J Med. 2014;127:939-53.

Description: We conducted a MEDLINE/EMBASE/CENTRAL search for randomized trials evaluating β-blockers in myocardial infarction enrolling at least 100 patients. The primary outcome was all-cause mortality. Analysis was performed stratifying trials into reperfusion-era (> 50% undergoing reperfusion or receiving aspirin/statin) or pre-reperfusion-era trials.

Results: Sixty trials with 102,003 patients satisfied the inclusion criteria. In the acute myocardial infarction trials, a significant interaction (Pinteraction = .02) was noted such that β-blockers reduced mortality in the pre-reperfusion (incident rate ratio [IRR] 0.86; 95% confidence interval [CI], 0.79-0.94) but not in the reperfusion era (IRR 0.98; 95% CI, 0.92-1.05). In the pre-reperfusion era, β-blockers reduced cardiovascular mortality (IRR 0.87; 95% CI, 0.78-0.98), myocardial infarction (IRR 0.78; 95% CI, 0.62-0.97), and angina (IRR 0.88; 95% CI, 0.82-0.95), with no difference for other outcomes. In the reperfusion era, β-blockers reduced myocardial infarction (IRR 0.72; 95% CI, 0.62-0.83) (number needed to treat to benefit [NNTB] = 209) and angina (IRR 0.80; 95% CI, 0.65-0.98) (NNTB = 26) at the expense of increase in heart failure (IRR 1.10; 95% CI, 1.05-1.16) (number needed to treat to harm [NNTH] = 79), cardiogenic shock (IRR 1.29; 95% CI, 1.18-1.41) (NNTH = 90), and drug discontinuation (IRR 1.64; 95% CI, 1.55-1.73), with no benefit for other outcomes. Benefits for recurrent myocardial infarction and angina in the reperfusion era appeared to be short term (30 days).

Conclusion: In contemporary practice of treatment of myocardial infarction, β-blockers have no mortality benefit but reduce recurrent myocardial infarction and angina (short-term) at the expense of increase in heart failure, cardiogenic shock, and drug discontinuation. The guideline authors should reconsider the strength of recommendations for β-blockers post myocardial infarction.

Bauters C, Lemesle G, Meurice T, Tricot O, de Groote P, Lamblin N. Prognostic impact of ß-blocker use in patients with stable coronary artery disease. Heart. 2014;100:1757-61.

Description: We analysed the data of 4184 outpatients included in a prospective cohort study on stable CAD. Two groups were formed based on ß-blocker use at enrollment. Two propensity score analyses were performed to control for differences in covariates: one with adjustment among the entire cohort, and the other with propensity score matching. The outcome variable was cardiovascular mortality after a 2-year follow-up.

Results: There were 3320 patients with ß-blocker use. Younger age, hypertension, diabetes, prior myocardial infarction, multivessel CAD, prior coronary revascularisation, prior stroke, prior hospitalisation for heart failure and a low LVEF were associated with ß-blocker use. Clinical follow-up data were obtained for 4149 patients (99.2%). When adjusted on propensity score, ß-blocker use was associated with a HR for cardiovascular mortality of 0.64 (0.42-0.98) in the whole cohort (p=0.04). After one-to-one propensity score matching, both groups (n=839 in each group) were well matched on covariates. The cardiovascular mortality rate in the propensity-matched cohort was significantly lower in patients with ß-blocker use with a HR of 0.43 (0.22-0.82) (p=0.011). Non-cardiovascular mortality was similar in both groups. These results were consistent across different subgroups.

Conclusions: In this observational study of patients with stable CAD, the use of ß-blockers was associated with a lower risk of cardiovascular mortality.

Impact on conclusions of systematic review:

These observational study does further support the recommendations and level of evidence ratings for this process of care. However, the other reviews do demonstrate that doing further guideline updates, careful review will need to be considered as to the utility of including this beta blocker measure after MI and including this subcomponent in the composite 0965 application.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1529

Measure Title: Beta Blocker at Discharge for ICD Implant Patients with LVSD

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: 965 Patients with an ICD implant who receive ACE-I/ARB and beta blocker therapy at discharge

Date of Submission: Click here to enter a date

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or

Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Beta-blocker therapy for patients with LVSD receiving an ICD

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

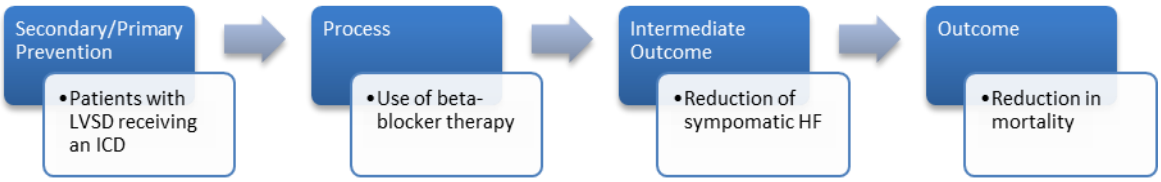
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



Beta-blockers reduce morbidity, mortality, and hospitalizations for patients with heart failure and left ventricular systolic dysfunction.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

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.

http://content.onlinejacc.org/article.aspx?articleid=1695825

Smith SC Jr., Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, Gibbons RJ, Grundy SM, Hiratzka LF, Jones DW, Lloyd-Jones DM, Minissian M, Mosca L, Peterson ED, Sacco RL, Spertus J, Stein JH, Taubert KA. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation. 2011: published online before print November 3, 2011, 10.1161/CIR.0b013e318235eb4d.

http://content.onlinejacc.org/article.aspx?articleid=1147807

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

2013 ACCF/AHA Guideline for the Management of Heart Failure (p. e169-170, 176, 195)

Stages of Heart Failure:

Stage A: At high risk for HF, but without structural heart disease or symptoms of Failure

Stage B: Structural heart disease, but without signs or symptoms of HF

Stage C: Structural heart disease with prior or current symptoms of HF

Stage D: Refractory HF requiring specialized interventions

p. e169

Stage B:

1. In all patients with a recent or remote history of MI or ACS and reduced EF, evidence-based beta blockers should be used to reduce mortality. **Class I: Level of Evidence: B**
2. Beta blockers should be used in all patients with a reduced EF to prevent symptomatic HF, even if they do not have a history of MI. **Class I: Level of Evidence: C**

p. e176:

Stage C:

3. Use of 1 of the 3 beta blockers proven to reduce mortality (e.g., bisoprolol, carvedilol, and sustained-release metoprolol succinate) is recommended for all patients with current or prior symptoms of HFrEF, unless contraindicated, to reduce morbidity and mortality. **Class I: Level of Evidence: A**

AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update (p. 2435)

4. Beta-blocker therapy should be used in all patients with left ventricular systolic dysfunction (ejection fraction <40%) with heart failure or prior myocardial infarction, unless contraindicated. (Use should be limited to carvedilol, metoprolol succinate, or bisoprolol, which have been shown to reduce mortality.) **Class I: Level of Evidence: A**

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

Guideline Statement # (see 1a.4.2 above)	Class of Recommendation/Level of Evidence (for definitions see 1a.4.4 below)

1	Class Ib
2	Class Ic
3	Class Ia
4	Class Ib
5	Class Ib
6	Class Ia

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Class of Recommendation (COR) is an estimate of the size of the treatment effect considering risks versus benefits in addition to evidence and/or agreement that a given treatment or procedure is or is not useful/effective or in some situations may cause harm.

Class I: Procedure/Treatment should be performed/administered

Class IIa: It is reasonable to perform procedure/administer treatment

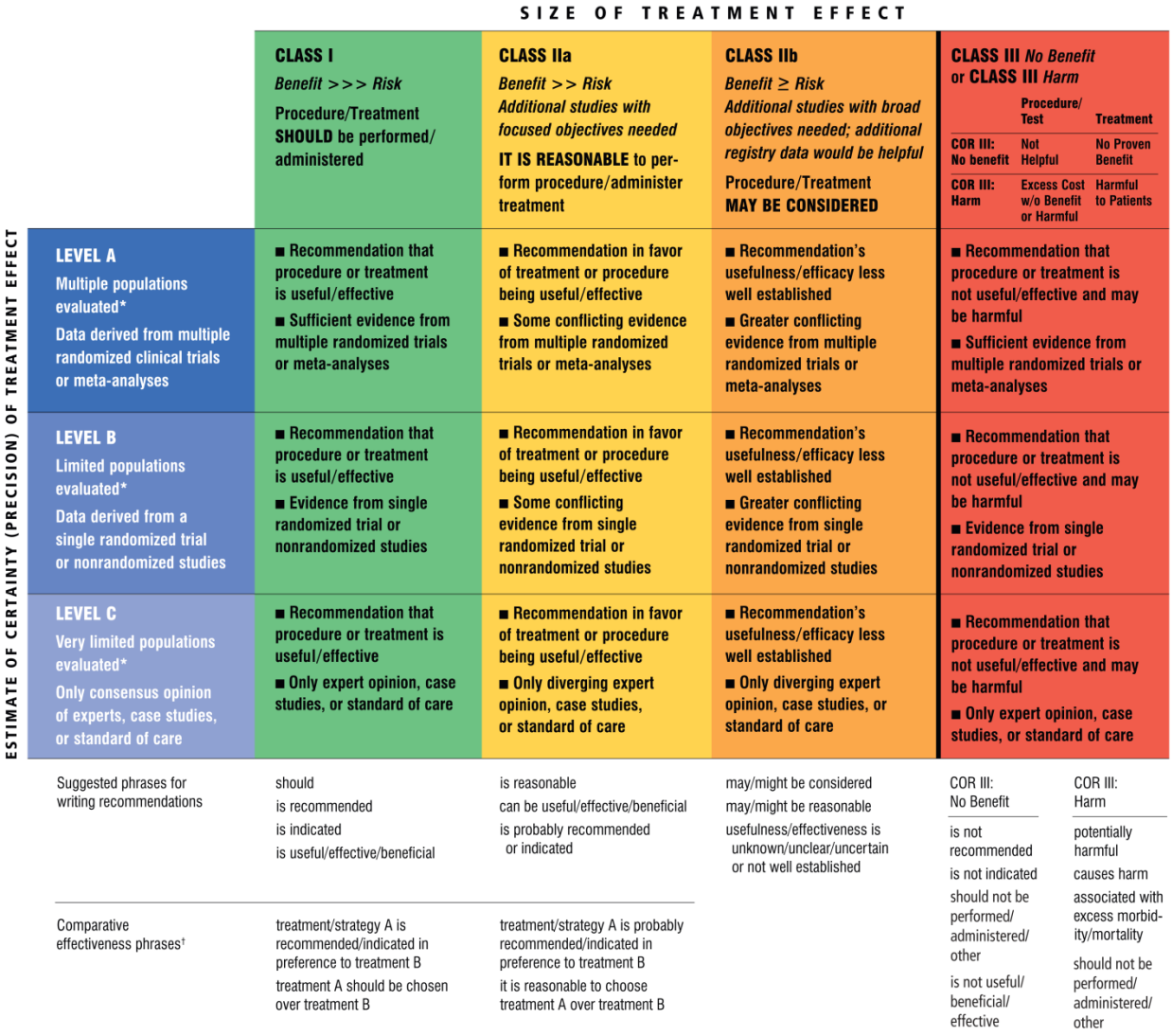
Class IIb: Procedure/Treatment may be considered

Class III: No benefit (Not helpful or No proven benefit)

Class III: Harm (Excess cost w/o benefit or Harmful to patients)

Specific COR definitions are included in Table 1 below.

Table 1. Applying Classification of Recommendation and Level of Evidence



Note: A recommendation with Level of Evidence B or C does not imply that the recommendation is weak. Many important clinical questions addressed in the guidelines do not lend themselves to clinical trials. Although randomized trials are unavailable, there may be a very clear clinical consensus that a particular test or therapy is useful or effective. *Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use. †For comparative effectiveness recommendations (Class I and IIa; Level of Evidence A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

ACCF/AHA Task Force on Practice Guidelines. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines. American College of Cardiology Foundation and American Heart Association, Inc. Cardiosource.com. 2010. Available at:

http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf and

http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

2013 ACCF/AHA Guideline for the Management of Heart Failure

This guideline covers multiple management issues for the adult patient with Heart Failure (HF) including the guideline-directed medical therapy (GDMT) such as beta-blocker therapy.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

An overall grade for the quality of evidence was not assigned. Rather, the quality of a study (or set of studies) supporting a recommendation was graded on an estimate of the certainty or precision of the treatment effect (see 1a.4.3).

Recommendations used to support this measure have a:

- Level of Evidence A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation.

OR

- Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies. References used to determine level of evidence must be provided and cited with the recommendation.

OR

- Level of Evidence C: Consensus opinion of experts, case studies, or standard of care.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Level of Evidence (LOE) is an estimate of the certainty or precision of the treatment effect.

Level of Evidence A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence C: Consensus opinion of experts, case studies, or standard of care.

Specific LOE definitions are included in Table 1 in 1a.4.4.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: An extensive evidence review was conducted through October 2011 and includes selected other references through April 2013.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The body of evidence supporting the recommendations on beta-blocker therapy for patients with LVSD includes:

7 randomized controlled trials

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

All but one of the recommendations for this process is rated as Level of Evidence A or B, meaning that the data was derived from one or more RCTs or meta-analyses. Additional information on the overall quality of evidence across the RCTs is not provided.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

p. e170:

CAD is a major risk factor for the development of HF and a key target for prevention of HF. The 5-year risk of developing HF after acute MI is 7% and 12% for men and women, respectively; for men and women between the ages of 40 and 69 and those >70 years of age, the risk is 22% and 25%, respectively. Current evidence supports the use of ACE inhibitors and (to a lower level of evidence) beta-blocker therapy to impede maladaptive LV remodeling in patients with stage B HF and low LVEF to improve mortality and morbidity.

In 1 study, losartan reduced adverse outcomes in a population with hypertension, and in another study of patients post-MI with low LVEF, valsartan was equivalent to captopril. Data with beta blockers are less convincing in a population with known CAD, although in 1 trial carvedilol therapy in patients with stage B and low LVEF was associated with a 31% relative risk reduction in adverse long-term outcomes. In patients with previously established structural heart disease, the administration of agents known to have negative inotropic properties such as non-dihydropyridine calcium channel blockers and certain antiarrhythmics should be avoided.

p. e176:

Three beta blockers have been shown to be effective in reducing the risk of death in patients with chronic HFrEF: bisoprolol and sustained-release metoprolol (succinate), which selectively block beta-1–receptors; and carvedilol, which blocks alpha-1–, beta-1–, and beta-2–receptors. Positive findings with these 3 agents, however, should not be considered a beta-blocker class effect. Bucindolol lacked uniform effectiveness across different populations, and short-acting metoprolol tartrate was less effective in HF clinical trials. Beta-1 selective blocker nebivolol demonstrated a modest reduction in the primary endpoint of all-cause mortality or cardiovascular hospitalization but did not affect mortality alone in an elderly population that included patients with HFpEF.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

2013 ACCF/AHA Guideline for the Management of Heart Failure

p. e177:

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.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

One RCT, one prospective cohort study, and two meta-analyses were published after the publication of the 2013 ACCF/AHA Guideline for the Management of Heart Failure.

Note: Text below for description and results is verbatim from the article abstract.

Peck KY, Lim YZ, Hopper I, Krum H. Medical therapy versus implantable cardioverter-defibrillator in preventing sudden cardiac death in patients with left ventricular systolic dysfunction and heart failure: a meta-analysis of >35,000 patients. Int J Cardiol. 2014: 173(2): 197-2003.

Description: Our meta-analysis included trials of >100 patients with reduced left ventricular ejection fraction (LVEF), i.e.,<40%. Fourteen randomized controlled trials met the criteria for meta-analysis, 10 involving medical therapies (angiotensin receptor blockers [ARBs], mineralocorticoid receptor antagonists [MRAs], ivabradine, n3-polyunsaturated fatty acid [PUFA], ferric carboxymaltose and aliskiren) and four involving ICDs. Results were pooled using the Mantel-Haenszel random effects method.

Results: Drug therapy (*n* = 36,172) reduced the risk of SCD overall (risk ratio (RR) = 0.89, 95% confidence interval (CI) = 0.82–0.98, *p* = 0.02) when compared to placebo. MRAs alone were most effective in reducing SCD (*n*= 11,032, RR = 0.79 [0.68–0.91], *p* = 0.001). ICD insertion greatly reduced SCD (*n* = 4,269, RR = 0.39 [0.30–0.51], *p* < 0.00001) compared with placebo. The difference in treatment effect between the ICD and drug therapy was significant (*p* < 0.002), and between ICD and MRAs (*p* < 0.002).

Al-Gobari M, El Khatib C, Pillion F, Gueyffier F. β-Blockers for the prevention of sudden cardiac death in heart failure patients: a meta-analysis of randomized controlled trials. BMC Cardiovasc Disord. 2014: 13:52.

Description: We conducted a meta-analysis of all randomized controlled trials examining the use of beta-blockers vs. placebo/control for the prevention of SCD in heart failure patients. We identified 30 trials, which randomized 24,779 patients to beta-blocker or placebo/control. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed. Eligible studies had to be randomized controlled trials and provide information on the incidence of sudden cardiac death in heart failure patients. Additional inclusion criteria included: treatment for >30 days and follow-up ≥3 months. Studies of patients<18 years, randomization to beta-blocker vs. an angiotensin converting enzyme (without placebo) and/or beta-blocker in both arms were excluded from the analysis. Pre-specified outcomes of interest included SCD, cardiovascular death (CVD), and all-cause mortality and were analyzed according to intention-to-treat.

Results: We found that beta-blockers are effective in the prevention of SCD [OR 0.69; 95% CI, 0.62-0.77, P<0.00001], cardiovascular death (CVD) [OR 0.71; 95% CI, 0.64-0.79, P<0.00001], and all-cause mortality [OR 0.67; 95% CI, 0.59-0.76, P<0.00001]. Based on the study analysis, 43 patients must be treated with a beta-blocker to prevent one SCD, 26 patients to prevent one CVD and 21 patients to prevent all-cause mortality in one year.

Conclusion: Beta-blockers reduce the risk of sudden cardiac death (SCD) by 31%, cardiovascular death (CVD) by 29% and all-cause mortality by 33%. These results confirm the mortality benefits of these drugs and they should be recommended to all patients similar to those included in the trials.

Fiuzat M, Wojdyla D, Kitzman D, Fleg J, Keteyian SJ, Kraus WE, Pina IL, Whellan D, O'Connor CM. Relationship of beta-blocker dose with outcomes in ambulatory heart failure patients with systolic dysfunction: results from the HF-ACTION (Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training) trial. J AM Coll Cardiol. 2012;60:208-15.

Description: The HF-ACTION trial was a randomized, multicenter trial enrolling 2,331 ambulatory HF patients with systolic dysfunction (New York Heart Association functional class II to IV, left ventricular ejection fraction <0.35) randomized to exercise training versus usual care, with median follow-up of 2.5 years. The BB dose at baseline was standardized with carvedilol equivalents and analyzed as a continuous variable and by discrete dose groups. The relationship between BB dose and the primary endpoint of all-cause mortality or all-cause hospitalization and other cardiovascular secondary endpoints was determined before and after adjustment for variables significantly associated with outcomes in the HF-ACTION cohort.

Results: Ninety-five percent of patients were receiving a BB. There was a significant inverse relationship between BB dose and all-cause death or hospitalization but not other cardiovascular endpoints after adjustment for other predictors of outcome, with a linear benefit up to the 50-mg daily dose. There was a significant association between BB dose and change in peak VO(2) at 3 months. There was no increase in bradycardia with higher doses of BB.

Bauters C, Lemesle G, Meurice T, Tricot O, de Groote P, Lamblin N. Prognostic impact of ß-blocker use in patients with stable coronary artery disease. Heart. 2014;100:1757-61.

Description: We analysed the data of 4184 outpatients included in a prospective cohort study on stable CAD. Two groups were formed based on ß-blocker use at enrollment. Two propensity score analyses were performed to control for differences in covariates: one with adjustment among the entire cohort, and the other with propensity score matching. The outcome variable was cardiovascular mortality after a 2-year follow-up.

Results: There were 3320 patients with ß-blocker use. Younger age, hypertension, diabetes, prior myocardial infarction, multivessel CAD, prior coronary revascularisation, prior stroke, prior hospitalisation for heart failure and a low LVEF were associated with ß-blocker use. Clinical follow-up data were obtained for 4149 patients (99.2%). When adjusted on propensity score, ß-blocker use was associated with a HR for cardiovascular mortality of 0.64 (0.42-0.98) in the whole cohort (p=0.04). After one-to-one propensity score matching, both groups (n=839 in each group) were well matched on covariates. The cardiovascular mortality rate in the propensity-matched cohort was significantly lower in patients with ß-blocker use with a HR of 0.43 (0.22-0.82) (p=0.011). Non-cardiovascular mortality was similar in both groups. These results were consistent across different subgroups.

Conclusions: In this observational study of patients with stable CAD, the use of ß-blockers was associated with a lower risk of cardiovascular mortality.

Impact on conclusions of systematic review: These studies and meta-analyses further support the recommendations and level of evidence ratings for this process of care.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1522

Measure Title: ACE/ARB Therapy at Discharge for ICD Implant Patients with LVSD

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: 965 Patients with an ICD implant who receive ACE-I/ARB and beta blocker therapy at discharge

Date of Submission: Click here to enter a date

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or

Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: ACE/ARB therapy for patients with LVSD receiving an ICD

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

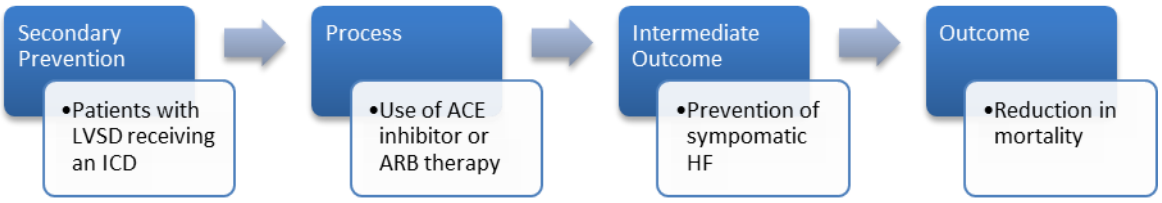
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



Angiotensin-converting enzyme (ACE) inhibitors and angiotensin-receptor antagonists/blockers (ARBs) reduce morbidity, mortality, and hospitalizations for patients with heart failure and left ventricular systolic dysfunction.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

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.

http://content.onlinejacc.org/article.aspx?articleid=1695825

Smith SC Jr., Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, Gibbons RJ, Grundy SM, Hiratzka LF, Jones DW, Lloyd-Jones DM, Minissian M, Mosca L, Peterson ED, Sacco RL, Spertus J, Stein JH, Taubert KA. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation. 2011: published online before print November 3, 2011, 10.1161/CIR.0b013e318235eb4d.

http://content.onlinejacc.org/article.aspx?articleid=1147807

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

2013 ACCF/AHA Guideline for the Management of Heart Failure (e169-170, 174-175, 195)

Stages of Heart Failure:

Stage A: At high risk for HF, but without structural heart disease or symptoms of Failure

Stage B: Structural heart disease, but without signs or symptoms of HF

Stage C: Structural heart disease with prior or current symptoms of HF

Stage D: Refractory HF requiring specialized interventions

Stage B recommendations (e169-170):

1. In all patients with a recent or remote history of MI or ACS and reduced EF, ACE inhibitors should be used to prevent symptomatic HF and reduce mortality. In patients intolerant of ACE inhibitors, ARBs are appropriate unless contraindicated. **Class I; Level of Evidence: A**
2. ACE inhibitors should be used in all patients with a reduced EF to prevent symptomatic HF, even if they do not have a history of MI. **Class I; Level of Evidence: A**

Stage C recommendations (e174-175):

3. ACE inhibitors are recommended in patients with HFrEF and current or prior symptoms, unless contraindicated, to reduce morbidity and mortality. **Class I; Level of Evidence: A**
4. ARBs are recommended in patients with HFrEF with current or prior symptoms who are ACE inhibitor intolerant, unless contraindicated, to reduce morbidity and mortality. **Class I; Level of Evidence: A**

AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update (p. 2435)

5. ACE inhibitors should be started and continued indefinitely in all patients with left ventricular ejection fraction <40% and in those with hypertension, diabetes, or chronic kidney disease, unless contraindicated. **Class I; Level of Evidence: A**

6. The use of ARBs is recommended in patients who have heart failure or who have had a myocardial infarction with left ventricular ejection fraction □40% and who are ACE-inhibitor intolerant. **Class I; Level of Evidence: A**

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Guideline Statement # (see 1a.4.2 above)	Class of Recommendation/Level of Evidence (for definitions see 1a.4.4 below)
1	Class Ia
2	Class Ia
3	Class Ia
4	Class Ia
5	Class Ib
6	Class Ia
7	Class Ia

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Class of Recommendation (COR) is an estimate of the size of the treatment effect considering risks versus benefits in addition to evidence and/or agreement that a given treatment or procedure is or is not useful/effective or in some situations may cause harm.

Class I: Procedure/Treatment should be performed/administered

Class IIa: It is reasonable to perform procedure/administer treatment

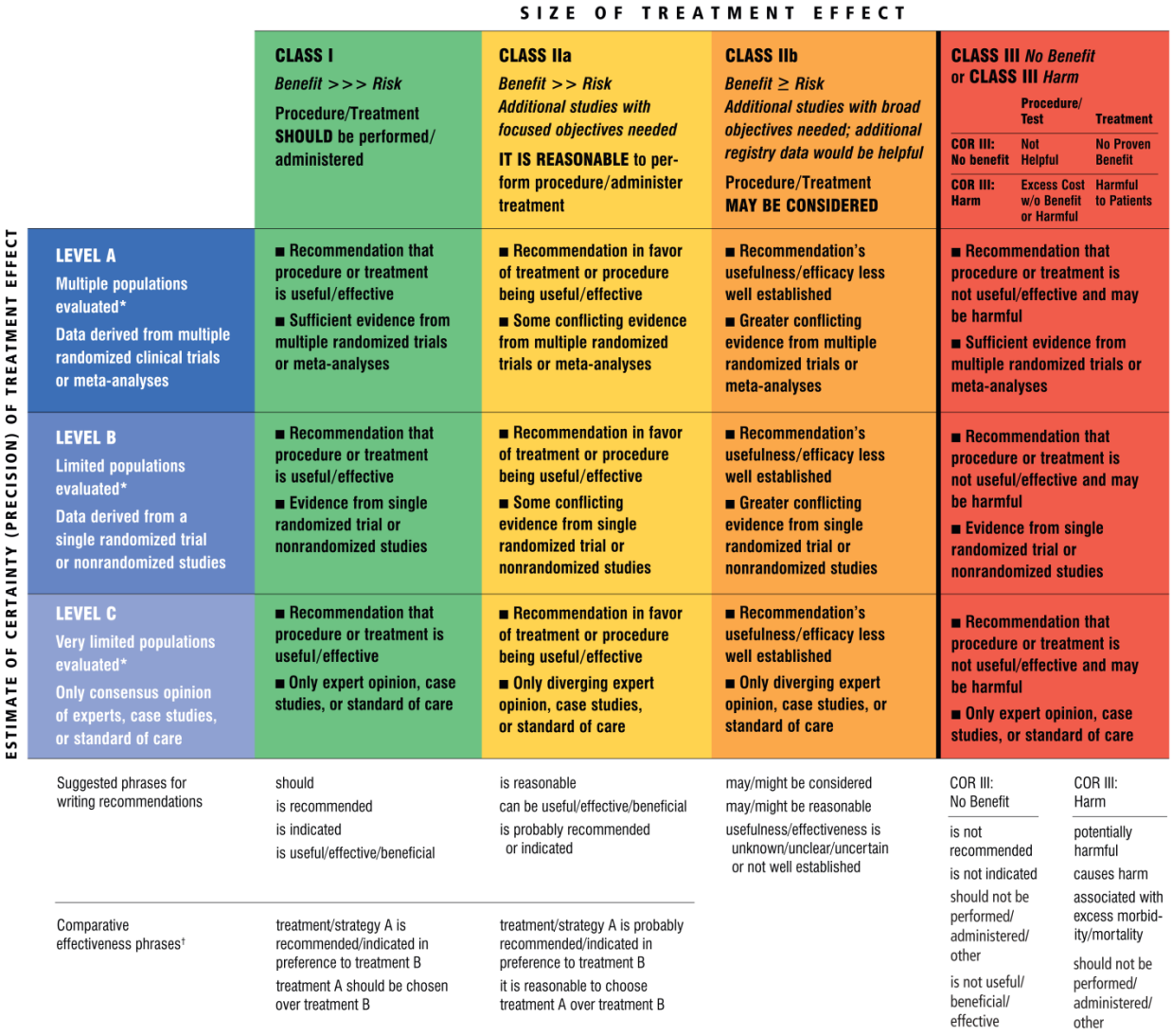
Class IIb: Procedure/Treatment may be considered

Class III: No benefit (Not helpful or No proven benefit)

Class III: Harm (Excess cost w/o benefit or Harmful to patients)

Specific COR definitions are included in Table 1 below.

Table 1. Applying Classification of Recommendation and Level of Evidence



Note: A recommendation with Level of Evidence B or C does not imply that the recommendation is weak. Many important clinical questions addressed in the guidelines do not lend themselves to clinical trials. Although randomized trials are unavailable, there may be a very clear clinical consensus that a particular test or therapy is useful or effective. *Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use. †For comparative effectiveness recommendations (Class I and IIa; Level of Evidence A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

ACCF/AHA Task Force on Practice Guidelines. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines. American College of Cardiology Foundation and American Heart Association, Inc. Cardiosource.com. 2010. Available at:

http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf and

http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

2013 ACCF/AHA Guideline for the Management of Heart Failure

This guideline covers multiple management issues for the adult patient with Heart Failure (HF) including the guideline-directed medical therapy (GDMT) such as ACE inhibitor or ARB therapies.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

An overall grade for the quality of evidence was not assigned. Rather, the quality of a study (or set of studies) supporting a recommendation was graded on an estimate of the certainty or precision of the treatment effect (see 1a.4.3).

Recommendations used to support this measure have a:

- Level of Evidence of A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation

OR

- Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies. References used to determine level of evidence must be provided and cited with the recommendation.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Level of Evidence (LOE) is an estimate of the certainty or precision of the treatment effect.

Level of Evidence A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence C: Consensus opinion of experts, case studies, or standard of care.

Specific LOE definitions are included in Table 1 in 1a.4.4.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: An extensive evidence review was conducted through October 2011 and includes selected other references through April 2013.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

The body of evidence supporting the recommendations on ACE/ARB therapy includes:

15 randomized controlled trials (RCTs)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

All but one of the recommendations for this process are rated as Level of Evidence A, meaning that the data was derived from multiple RCTs or meta-analyses. Additional information on the overall quality of evidence across the RCTs is not provided.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

p. e170:

CAD is a major risk factor for the development of HF and a key target for prevention of HF. The 5-year risk of developing HF after acute MI is 7% and 12% for men and women, respectively; for men and women between the ages of 40 and 69 and those >70 years of age, the risk is 22% and 25%, respectively. Current evidence supports the use of ACE inhibitors and (to a lower level of evidence) beta-blocker therapy to impede maladaptive LV remodeling in patients with stage B HF and low LVEF to improve mortality and morbidity. At 3-year follow-up, those patients treated with ACE inhibitors demonstrated combined endpoints of reduced hospitalization or death, a benefit that extended up to a 12-year follow-up. ARBs are reasonable alternatives to ACE inhibitors.

p. e175-176:

In several placebo-controlled studies, long-term therapy with ARBs produced hemodynamic, neurohormonal, and clinical effects consistent with those expected after interference with the renin-angiotensin system. Reduced hospitalization and mortality have been demonstrated. ACE inhibitors remain the first choice for inhibition of the renin-angiotensin system in systolic HF, but ARBs can now be considered a reasonable alternative.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

2013 ACCF/AHA Guideline for the Management of Heart Failure

p. e174-175:

The majority of the adverse reactions of ACE inhibitors can be attributed to the 2 principal pharmacological actions of these drugs: those related to angiotensin suppression and those related to kinin potentiation. Other types of adverse effects may also occur (e.g., rash and taste disturbances). Up to 20% of patients will experience an ACE inhibitor–induced cough. With the use of ACE inhibitors, particular care should be given to the patient's volume status, renal function, and concomitant medications. However, most HF patients (85% to 90%) can tolerate these drugs.

p. e176:

The risks of ARBs are attributed to suppression of angiotensin stimulation. These risks of hypotension, renal dysfunction, and hyperkalemia are greater when combined with another inhibitor of this neurohormonal axis, such as ACE inhibitors or aldosterone antagonists.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

One meta-analysis was published after the publication of the 2013 ACCF/AHA Guideline for the Management of Heart Failure.

Peck KY, Lim YZ, Hopper I, Krum H. Medical therapy versus implantable cardioverter-defibrillator in preventing sudden cardiac death in patients with left ventricular systolic dysfunction and heart failure: a meta-analysis of >35,000 patients. Int J Cardiol. 2014: 173(2): 197-2003.

Note: Text below for description and results is verbatim from the article abstract.

Description: Our meta-analysis included trials of >100 patients with reduced left ventricular ejection fraction (LVEF), i.e.,<40%. Fourteen randomized controlled trials met the criteria for meta-analysis, 10 involving medical therapies (angiotensin receptor blockers [ARBs], mineralocorticoid receptor antagonists [MRAs], ivabradine, n3-polyunsaturated fatty acid [PUFA], ferric carboxymaltose and aliskiren) and four involving ICDs. Results were pooled using the Mantel-Haenszel random effects method.

Results: Drug therapy (n = 36,172) reduced the risk of SCD overall (risk ratio (RR) = 0.89, 95% confidence interval (CI) = 0.82–0.98, p = 0.02) when compared to placebo. MRAs alone were most effective in reducing SCD (n= 11,032, RR = 0.79 [0.68–0.91], p = 0.001). ICD insertion greatly reduced SCD (n = 4,269, RR = 0.39 [0.30–0.51], p < 0.00001) compared with placebo. The difference in treatment effect between the ICD and drug therapy was significant (p < 0.002), and between ICD and MRAs (p < 0.002).

Impact on conclusions of systematic review: This meta-analysis further supports the recommendations and level of evidence ratings for this process of care.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

965_Composite_all_evidence_submission_061715_submitted.docx

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., the benefits or improvements in quality envisioned by use of this measure)*

This measure is intended to assess the extent to which eligible patients receive evidence-based medications that are indicated at hospital discharge following ICD placement. This measure focuses on processes of care that are supported by guidelines for optimal care for patients undergoing ICD placement.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

In the time period 2011-12, a total of 243,186 patients at 1552 hospitals were analyzed, and 195, 563 patients at 1606 hospitals in 2013-14. The distribution of the hospital performance was as follows:

2011 - 2012
Hospitals: n=1552
Patients: n=243186
Mean: 74%
Std Deviation: 16%
Percentiles
90th: 91%
75th (Quartile 3): 84%
50th (Median): 76%
25th (Quartile 1): 67%
10th: 56%

2013 - 2014
Hospitals: n=1606
Patients: n=195563

Mean: 78%
Std Deviation: 17%
Percentiles
90th: 97%
75th (Quartile 3): 89%
50th (Median): 79%
25th (Quartile 1): 71%
10th: 59%

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.**
There is a demonstration for an opportunity for improvement based on the noted performance ranges.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
See testing supplement for details.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
None

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Frequently performed procedure, High resource use, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Optimal medical therapy is critical to ensure favorable patient outcomes following implantation of an implantable cardiac defibrillator (ICD) to prevent sudden cardiac death (SCD). Approximately 86 million American adults have 1 or more types of CVD. Over 30% of all deaths are related to CVD. Nearly 787,000 people in the U.S. died from heart disease, stroke and other cardiovascular diseases in 2011 or about one of every three deaths in America. ACE inhibitors have been shown to decrease morbidity, mortality, and hospitalizations for patients with heart failure and left ventricular systolic dysfunction. The efficacy of ARB therapy has been strengthened by several large-scale prospective randomized clinical trials demonstrating reduction in mortality and hospitalization for heart failure among patients with heart failure and LVSD. Refer to evidence supplement for additional details. Long term beta blocker therapy for patients with left systolic ventricular dysfunction (LVSD) can improve symptoms of heart failure, improve patient clinical status, and reduce hospitalizations and mortality. Lastly, the benefits of beta blocker therapy in patients with prior myocardial infarction without contraindications have been established for a wide range of

patient groups. Because the majority of patients undergoing ICD implantation have LVSD and/or have had a prior myocardial infarction, a substantial proportion of this population has an indication for chronic ACE/ARB and/or beta blocker therapy to improve outcomes.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

American Heart Association. Heart disease and stroke statistics- 2014 update: A report of the American Heart Association. Circulation. 2014 Jan 21;129(3):e28-e292. doi: 10.1161/01.cir.0000441139.02102.80. Epub 2013 Dec 18.Accessed April 17, 2015.

Bonow RO, Bennett S, Casey DE, Jr., et al. ACC/AHA clinical performance measures for adults with chronic heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures (Writing Committee to Develop Heart Failure Clinical Performance Measures) endorsed by the Heart Failure Society of America. J Am Coll Cardiol. 2005;46:1144-78.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

1d. Composite Quality Construct and Rationale

1d.1. **A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
  - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient); or
  - any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient).

1d.1. Please identify the composite measure construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)

1d.2. **Describe the quality construct, including:**

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

This measure focuses on processes of care that recommended for optimal care for patients following ICD implantation. Each component of the measure has been shown in randomized clinical trials to impact clinical outcomes and represents a Class 1 guideline indication for the care of patients with left ventricular systolic dysfunction or prior myocardial infarction. Combining the individual process measures into a single composite provides patients, physicians, and hospitals with a perspective of the overall quality of medical therapy provided to patients undergoing ICD implantation. Hospitals with a gap in performance can investigate the individual components of the measure to identify specific opportunities for improvement. The content validity of this measure has been achieved by virtue of their consistency with strong guideline recommendations and the expertise of the individuals who developed this measure.

In addition, we conducted empiric analyses examining the association between performance on the composite measure and clinical outcomes including readmission and mortality at 6 months following device implantation (see testing supplement for

detailed results). We found that patients who were discharged on appropriate medical therapy were less likely to experience adverse outcomes compared with patients who were not discharged on appropriate medical therapy. Furthermore, fewer patients treated at high performing hospitals as determined by this composite experienced adverse outcomes compared with those treated at low performing hospitals

1d.3. **Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.**

This measure is intended to assess the extent to which eligible patients receive evidence-based medications that are indicated at hospital discharge following ICD implantation.

Composite performance measures have a variety of uses.

Data reduction. A large and growing array of individual indicators makes it possible for users to become overloaded with data. A composite measure reduces the information burden by distilling the available indicators into a simple summary.

Scope expansion. The information in a composite measure is highly condensed, making it feasible to track a broader range of metrics than would be possible otherwise. Composite measures have been described as a tool for making provider assessments more comprehensive

Provider performance valuation. Performance indicators are used for various decisions about providers, including the allocation of pay-for-performance incentives, designation of preferred provider status, and assignment of letter grades and star rating categories. If a decision is to be based on multiple indicators instead of a single indicator, a method of translating several variables into a single decision is needed. Composite measures serve this function by assigning providers to 1 position on a scale of

better-to-worse performance.

Given all these uses, NCDR believes that while we will continue to report these measures at the individual level there is a distinctive value of having a composite measure endorsed at NQF.

1d.4. **Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**

Each of the components of this measure address appropriate medication prescribing at discharge for ICD patients.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Cardiovascular

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

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

N/A

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** icd_v2_datadictionary_codersdictionary_2-1-635246241637392049.pdf

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
No change since last endorsement.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who receive ACE/ARB and Beta blockers for which they are eligible.
1. ACE/ARB prescribed at discharge (if eligible for ACE/ARB as described in denominator)
AND
2. Beta blockers prescribed at discharge (if eligible for beta blockers as described in denominator)

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
1 year

S.6. **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, 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.
If eligible for beta blocker and given, then code "Yes"
If eligible for beta blocker and not given, then code "No, not given"

If eligible for ACE/ARB and given, then code then "Yes"
If eligible for ACE/ARB and not given, then code "No, not given"

If any "No, not given" present, then performance not met. Else, performance met.

Note: Contraindicated and those participating in blinded studies are also considered as exceptions and performance met.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients with an ICD implant surviving hospitalization who are eligible to receive any one of the two medication classes:
1) Eligiblility for ACE/ARB: Patients who have an ejection fraction (EF) of <40% AND do not have a documented contraindication to ACE/ARB documented
OR
2) Eligibility for beta blockers: Patients who do not have a documented contraindication to beta blocker therapy and have

either:

a. EF of <40% OR

b. a previous myocardial infarction (MI)

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

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

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Discharge status of expired; not eligible for either ACE/ARB or beta blockers

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
NCDR makes a clear distinction between absolute "Exclusions" (e.g., death, transfer) and relative "Exceptions", (e.g., contraindications). While patients with exclusions are always automatically removed from the denominator and numerator, exceptions allow clinicians the opportunity to identify an intervention/process/medication as not clinically indicated based on the unique patient scenario.
Each of the two medications incorporated into this composite may be coded as Yes (medication prescribed), No (medication not prescribed), Blinded (pt. involved in a clinical trial, medication type unavailable for data entry), and Contraindicated (used to capture many of the medical exceptions used in this measure).

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**

Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
1) Remove patients whose discharge status is expired
2) Check if given patient is eligible for 1 of the 2 medication therapies.
3) If eligible for at least 1 medication, then keep this patient.
4) If not eligible for any of the 2 medications, then patient is removed from eligibility.

If eligible for ACE/ARB and given, then code "Yes"
If eligible for ACE/ARB and not given, then code "No, not given"
If eligible for ACE/ARB but contraindicated, then code "contraindicated/blinded"

If eligible for Beta Blocker and given, then code then "Yes"
If eligible for Beta Blocker and not given, then code "No, not given"
If eligible for Beta Blocker but contraindicated, then code "contraindicated/blinded"

5) If any "No, not given" present, then performance not met. Else, performance met.

Although ineligible cases are removed from the denominator population for the performance calculation, the number of patients with valid exceptions should be calculated and reported along with performance rates to track variations in care and highlight possible areas of focus for QI.

If the patient does not meet the numerator and a valid exception is not present, this case represents a quality failure.

Missing data defaults to "performance not met" This measure assumes that missing documentation on the process results in a failure of meeting an evidence based therapy.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)

Required for Composites and PRO-PMs.

Missing data defaults to "performance not met" This measure assumes that missing documentation on the process results in a failure of meeting a evidence based therapy.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
National Cardiovascular Data Registry (NCDR) ICD Registry

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
ACC_NCDR_ICD_Composite_Meds_Testing_Supplement__061715_submitted-635705722018946580.docx

NATIONAL QUALITY FORUM—Composite Measure Testing (subcriteria 2a2, 2b2-2b6, 2d)

Composite Measure Title: Discharge Medications (ACE/ARB and beta blockers) in Eligible ICD Implant Patients
Date of Submission: Click here to enter a date
Composite Construction:
☐Two or more individual performance measure scores combined into one score
☐**All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)**
☐Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

Instructions: Please contact NQF staff before you begin.

- If a component measure is submitted as an individual performance measure, the non-composite measure testing form must also be completed and attached to the individual measure submission.
- 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 <u>all</u> composite measures, sections 1, 2a2, 2b2, 2b3, 2b5, and 2d must be completed.**
- **For composites with <u>outcome and resource use</u> measures**, section **2b4** also must be completed.
- If specified for **<u>multiple data sources/sets of specifications</u>** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2), validity (2b2-2b6), and composites (2d) 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.

<u>Note</u>: The information provided in this form is intended to aid the Steering 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.

2b2. 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.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

Composite 2d. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

1) the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and
2) the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible; and
3) the extent of missing data and how the specified handling of missing data minimizes bias (i.e., achieves scores that are an accurate reflection of quality).

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.

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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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 different components in the composite, indicate the component after the checkbox.*)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ **clinical database/registry**	☐ **clinical database/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*).

We propose to use the National Cardiovascular Data Registry for ICD Registry. This is a national quality improvement registry used in >1700 US hospitals. Some states and healthcare systems mandate participation, and participation is required as a condition for hospital reimbursement for Medicare beneficiaries receiving ICD therapy for primary prevention of sudden death. Rigorous quality standards are applied to the data and both quarterly and performance reports are generated for participating centers to track and improve their performance.

1.3. What are the dates of the data used in testing? Click here to enter date range
We have chosen to use different datasets to provide support for different aspects of the proposed measure.
Assessment of item-level reliability through the Audit Program: 01/2010-12/2010
All other forms of reliability testing: Jan 2013-Jun 2014
Hospital information about the Safety Net Hospital and %Medicaid are derived from AHA 2010 data.

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item*	**Measure Tested at Level of:**
☐individual clinician	☐individual clinician
☐group/practice	☐group/practice

☐**hospital/facility/agency**	☐**hospital/facility/agency**
☐health plan	☐health plan
☐other: Click here to describe	☐other: Click here to describe

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

For all the descriptive statistics for this measure except auditing:

Number of the measured entities (hospitals): 1,606

<u>Assessment of item-level reliability through the Audit Program:</u>

To assess inter-rater reliability of the extracted data elements that comprise this measure, data from 25 participating hospitals were reviewed by an independent contractor hired by ACCF.

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

The number of patients varies by testing type.

For all the descriptive statistics for this measure except auditing we used data submitted to the ICD Registry between January 2013 and June 2014. Note this reflects all data from all centers that met data quality standards irrespective of the case volume of participating hospitals. When we present information about hospital performance

Selected Characteristics by Calendar Year

Description	Total		Year			
			Jan – Dec 2013		Jan – Jun 2014	
	#	%	#	%	#	%
ALL	**195563**	**100.00**	**131193**	**100.00**	**64370**	**100.00**
Age>=65						
No	**70743**	**36.17**	**47084**	**35.89**	**23659**	**36.75**
Yes	**124820**	**63.83**	**84109**	**64.11**	**40711**	**63.25**
Female						
No	**145765**	**74.54**	**97732**	**74.49**	**48033**	**74.62**
Yes	**49798**	**25.46**	**33461**	**25.51**	**16337**	**25.38**
RACE						
Hispanic	**11268**	**5.76**	**7541**	**5.75**	**3727**	**5.79**
White non-hispanic	**152042**	**77.75**	**102370**	**78.03**	**49672**	**77.17**
Black non-Hispanic	**27925**	**14.28**	**18421**	**14.04**	**9504**	**14.76**
Other	**4328**	**2.21**	**2861**	**2.18**	**1467**	**2.28**
Safety Net Hospital*						
Unknown	**2342**	**1.20**	**1588**	**1.21**	**754**	**1.17**
No	**164694**	**84.22**	**110503**	**84.23**	**54191**	**84.19**
Yes	**28527**	**14.59**	**19102**	**14.56**	**9425**	**14.64**
Hospital % Non-White						
Q1 (0.00% to 3.16%)	**27378**	**14.00**	**18254**	**13.91**	**9124**	**14.17**

Q2 (>3.16% to 10.57%)	**56591**	**28.94**	**37975**	**28.95**	**18616**	**28.92**
Q3 (>10.57% to 24.12%)	**64411**	**32.94**	**43568**	**33.21**	**20843**	**32.38**
Q4 (>24.12%)	**47183**	**24.13**	**31396**	**23.93**	**15787**	**24.53**
Hospital % Medicaid*						
Unknown	**2342**	**1.20**	**1588**	**1.21**	**754**	**1.17**
Q1 (0.00% to 12.70%)	**50024**	**25.58**	**33968**	**25.89**	**16056**	**24.94**
Q2 (>12.70% to 18.41%)	**52577**	**26.88**	**34940**	**26.63**	**17637**	**27.40**
Q3 (>18.41% to 22.72%)	**49841**	**25.49**	**33299**	**25.38**	**16542**	**25.70**
Q4 (>22.72%)	**40779**	**20.85**	**27398**	**20.88**	**13381**	**20.79**
Met the Composite Measure						
No	**36242**	**18.53**	**24699**	**18.83**	**11543**	**17.93**
Yes	**159321**	**81.47**	**106494**	**81.17**	**52827**	**82.07**

* Hospital information about the Safety Net Hospital and %Medicaid are derived from AHA 2010 data.

Assessment of item-level reliability through the Audit Program:
To assess inter-rater reliability of the extracted data elements that comprise this measure, we reviewed 627 patients.

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

There are different time periods and different descriptive statistics as noted in previous sections. The datasets, dates, number of measured entities, and number of admissions used in each type of testing are as follows:

For reliability testing (Section 2a2) using audit data: 01/2010 – 12/2010
For the split sample testing: 01/2013 – 06/2014

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We do not currently collect many of the SDS variables examples listed above. However, we do collect data on race as well as insurance type.

2a2. RELIABILITY TESTING
2a2.1. What level of reliability testing was conducted?
Note*: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*
☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. Describe the method of reliability testing and what it tests (*describe the steps—do not*

just name a method; what type of error does it test; what statistical analysis was used)

Split Sample Methodology

For the performance rates and disparities data, raw rates were calculated and a correlation coefficient was computed.

Assessment of item-level reliability through the Audit Program:

To assess inter-rater reliability of the extracted data elements that comprise this measure, 627 patients at 25 hospitals were reviewed by an independent contractor hired by ACCF.

Assessment of item-level reliability through the Audit Program:

The NCDR Data Quality Program ensures that data submitted to the NCDR are collected completely and in a valid manner. The NCDR Data Quality Program consists of 3 main components: data completeness, consistency, and accuracy. Completeness focuses on the proportion of missing data within fields, whereas consistency determines the extent to which logically related fields contain values consistent with other fields. Accuracy characterizes the agreement between registry data and the contents of original charts from the hospitals submitting data. Before entering the Enterprise Data Warehouse (EDW), all submissions are scored for file integrity and data completeness, receiving 1 of 3 scores that are transmitted back to facilities using a color-coding scheme. A "red light" means that a submission has failed because of file integrity problems such as excessive missing data and internally inconsistent data. Such data are not processed or loaded into the EDW. A "yellow light" status means that a submission has passed the integrity checks but failed in completeness according to predetermined thresholds. Such data are processed and loaded into the EDW but are not included in any registry aggregate computations until corrected. Facilities are notified about data submission problems and provided an opportunity to resubmit data. Finally, a "green light" means that a submission has passed all integrity and quality checks. Such submissions are loaded to the EDW. After passing the DQR, data are loaded into a common EDW that houses data from all registries and included for all registry aggregate computations. In a secondary transaction process, data are loaded into registry-specific, dimensionally modeled data marts.
A summary of the Program is noted under Table 1.

Table 1. Data Quality Program Overview

Methodology	• Nationwide program (i.e., all submitting participants in the United States) • Review of data submitted the previous year • Review of a subset of data elements that can rotate each year • Remote review of data combined with couple of onsite visit • Onsite visits are targeted based on the Data Outlier Program • Random selection of sites and records • Blinded data abstraction from medical charts • Inter-rater Reliability Assessment conducted to validate the audit findings • Adjudication step for participant to refute audit findings
Scope	• Review of hospital's medical records for related episodes of care • Assessment of complete submission (Comparison of two lists : hospital

	list of cases with specific billing codes versus NCDR submitted records)
Criteria for selecting sites/records	Remote audit : • Sites passing their quarterly DQR for 2 quarters within audited year • Sites submitting at least the number of records/sites being reviewed Onsite audit • Sites identified with an outlier and not contacted with the data outlier program
Scoring	NCDR uses a grading system for identifying the amount of agreement or matching between the data captured during the medical record review and data submitted to the NCDR.

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

Assessment of item-level reliability through the Audit Program:

CE #	field_Name	agreement rate	_KAPPA_	L_KAPPA	U_KAPPA	N levels
4170	Prior MI	0.815920398	0.60411	0.54118	0.66704	3
5000	LVEF Assessed	0.797678275	0.327125	0.245208	0.409041	3
6005	Procedure Type	0.978441128	0.955269	0.931311	0.979227	4
9045	ACE Inhibitor (Any)	0.883913765	0.755579	0.707206	0.803952	4
9100	ARB (Any)	0.922056385	0.729868	0.660493	0.799242	4
9110	Beta Blocker (Any)	0.933665008	0.658258	0.568486	0.748029	5

Assessment of item-level reliability through the Audit Program:

NCDR's Data Quality Program rotates the review of all the variables in the registry. ICD has over 300 elements that are reviewed on a 3 year rotating cycle. The elements required for this measure will be reviewed during the upcoming audit process. NCDR staff can provide kappa scores and percentage agreement scores upon completion of the cycle.

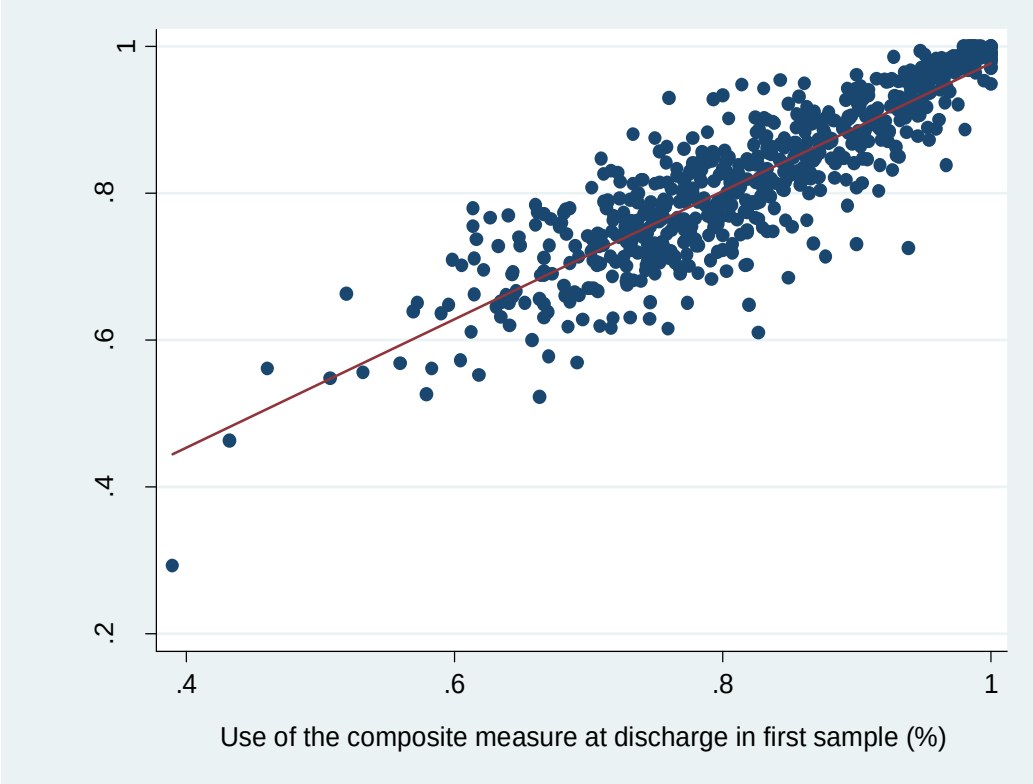
Split Sample Methodology:

Distribution of hospital performance on the composite measure within random split samples (minimum 50 cases in each sample)

Description	Randomly Split Samples	
	First (RAND=1) DCM	Second (RAND=0) DCM
N	707	684
Mean	0.8178	0.8200
Std Deviation	0.1089	0.1090
100% Max	1.0000	1.0000
75% Q3	0.9020	0.9087
50% Median	0.8199	0.8203
25% Q1	0.7414	0.7434

To evaluate the reliability of the measure, we randomly split the study cohort over the two year period (Jan 2013 to Jun 2014 combined) into two samples and restricted the cohort to hospitals that had a minimum of 50 cases in each split sample.

Results of the split sample testing are provided below. The 2 split samples were calculated during the same timeframe to avoid the potential for changes in hospital performance over time. After splitting the cohort into two random samples, we compared measure scores calculated at hospitals with at least 50 cases in both random samples. Of note, slightly less than half of participating hospitals met this volume threshold, and a few hospitals had more than 50 cases in one random sample but fewer than 50 in the other. The distribution of hospital performance was similar in the two samples (figure below), and there was an extremely high correlation between hospital performances assessed in the two samples (r 0.87949)



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

Assessment of item-level reliability through the Audit Program:

These kappa scores were calculated with a 95% CI. By convention, a kappa > .70 is considered acceptable inter-rater reliability (Landis 1977). We used the scale below for our analysis.

0: No better than chance
0.01-0.20: Slight
0.21-0.40: Fair
0.41-0.60: Moderate
0.61-0.80: Substantial
0.81-1.0: Almost perfect

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

The kappa score for all medication elements demonstrate substantial or almost perfect reliability. Some of the measure elements have justifiable reasons for a lower kappa and percentage agreement scores. The element "LVEF Assessed" is not always known. Moreover, there are multiple data elements at different times during hospitalization. Therefore, it is difficult to assess which

score is the correct score. Nevertheless, this element is actively discussed on monthly registry site manager calls and NCDR' s educational annual conference.

Split Sample Methodology

The figure above shows the scatterplot of the distribution of hospital performance for ICD composite measure at discharge when assessed in randomly split samples. Overall hospital performance in one random sample was strongly correlated with hospital performance in the other split sample (r=0.87949), which is consistent with a highly reliable measure.

2b2. VALIDITY TESTING

Note*: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

2b2.1. What level of validity testing was conducted?

☐**Composite performance measure score**

☐**Empirical validity testing**

☐**Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

☐**Systematic assessment of content validity**

☐**Validity testing for component measures** *(check all that apply)*

Note*: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☐**Endorsed (or submitted) as individual performance measures**

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

☐**Empirical validity testing of the component measure score(s)**

☐**Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level 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)*

Systematic assessment of content validity:

Content validity of this process was achieved by the specialized expertise of those individuals who developed this measure as well as the structured discussions that the group conducted. For this particular topic those individuals who were involved in identifying the key attributes and variables for this process measure were leaders and experts in the field of electrophysiology. Serial phone

calls were held to both define the eligible population and given process. These clinical leaders are noted below.

NCDR Clinical Measures workgroup ensured the measure demonstrated an opportunity for improvement, had strong clinical evidence, and was a reliable and valid measure. These members included Drs. Jeptha Curtis (Chair), Frederick Masoudi, John Rumsfeld, Mark Kremers, and Matthew Reynolds.

NCDR Scientific Quality and Oversight Committee—a committee that served as the primary resource for crosscutting scientific and quality of care methodological issues. These members included Drs. Frederick Masoudi (Chair) , David Malenka, Thomas Tsai, Matthew Reynolds, David Shahian, John Windle, Fred Resnic, John Moore, Deepak Bhatt, James Tcheng, Jeptha Curtis, Paul Chan, Matthew Roe, and John Rumsfeld.

Lastly the 16 member NCDR Management Board and 31member ACCF Board of Trustees reviewed and approved these measures for submission to NQF.

Evidence:

ACE/ARB

ACE inhibitors reduce morbidity, mortality, and hospitalizations for patients with heart failure and left ventricular systolic dysfunction. The efficacy of ARB therapy has been strengthened by several large-scale prospective randomized clinical trials demonstrating lower rates of death and heart failure hospitalization among patients with heart failure and LVSD. Consensus clinical guidelines include strong recommendations for ACE inhibitors for all patients with HF due to LV systolic dysfunction unless they have a contraindication to their use or have been shown to be unable to tolerate treatment with these drugs. ACE inhibitors remain the first choice for inhibition of the renin-angiotensin system in chronic HF, but ARBs are considered a reasonable alternative. Even if the patient has responded favorably to the diuretic, treatment with ACE inhibitor or ARBs should be initiated and maintained in patients who can tolerate them, because they have been shown to favorably influence the long-term prognosis of HF

Beta Blocker-MI

The benefits of beta blocker therapy in patients with prior myocardial infarction without contraindications have been established for a wide range of patient groups. The greatest benefits are seen in patients with the greatest baseline risk: those with impaired ventricular function or ventricular arrhythmias and those who do not undergo reperfusion. The benefits of beta-blocker therapy for secondary prevention are well established.

Beta Blocker-LVSD

Long term beta blocker therapy for patients with left systolic ventricular dysfunction (LVSD) can improve symptoms of heart failure, improve patient clinical status, and reduce hospitalizations and mortality.

All this research demonstrates that this measure contributes to improved intermediate outcomes and important outcomes such as reductions in hospitalizations and mortality rates.

Empiric assessment of content validity:

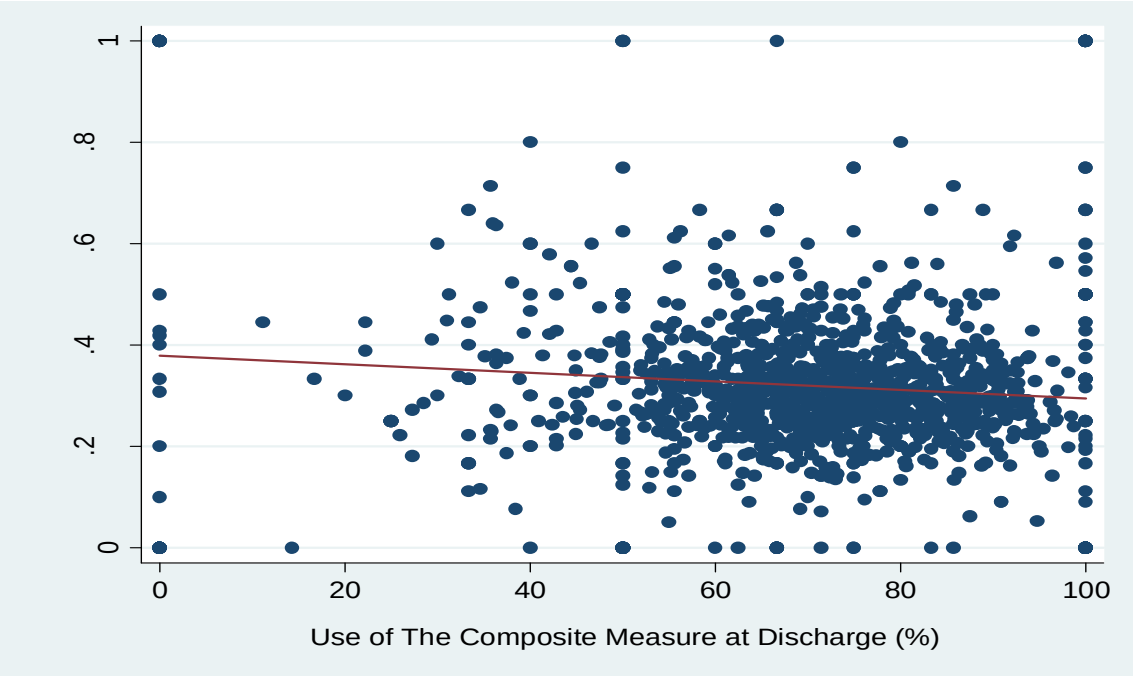
As noted in the measure application, we conducted empiric analyses to assess the association of patient and hospital performance on the composite measure with adverse outcomes, specifically mortality and readmission at 6 months following hospital discharge. To conduct these analyses we used a sample of patients for whom these outcomes were available. This consisted of 93971 Medicare fee-for-service patients at least 65 years of age who underwent ICD implantation in 2010 or 2011. Our outcomes of interest included all-cause mortality, all-cause readmission, and the combination of the 2 at 6 months following hospital discharge. We examined the proportion of patients who experienced these outcomes stratified by whether or not they were discharged on appropriate medical therapy. In addition, we conducted analyses at the hospital level examining the association between hospital-level performance on the measure and the combination of mortality or readmission at 6 months.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Patient-level results are shown below. Overall, a significantly smaller proportion of patients discharged on appropriate medical therapy died or were readmitted within 6 months of hospital discharge.

Description	Use of Medications				P
	No		Yes		
	#	%	#	%	
Composite Measure	**25217**		**68754**		
6 month mortality	2408	9.55	3720	5.41	<0.001
6 month readmission	8587	34.05	18643	27.12	<0.001
6 month mortality or readmission	9148	36.28	19504	28.37	<0.001

Hospital-level results are shown below. The figure shows the association between rate of death or readmission within 6 months of discharge, with the use of the composite measure at discharge. Hospital performance on the composite discharge medication measure were significantly correlated with the combined outcome of death or readmission such that patients treated at hospitals that performed better on the measure had better unadjusted outcomes that those treated at hospitals that performed worse on the measure (correlation coefficient (-0.0998), $p<0.001$).



2b2.4. 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?*

These findings support the validity of the composite discharge medication measure. At both the patient and hospital level, performance on the measure was associated with better outcomes at 6 months following discharge.

Threats to Validity:

Information Bias: There should be little concern for information bias since the care process is objective and there is a low likelihood of misreporting the given care process. Additionally, since there is only 1 data source that is used for NCDR inpatient registries thus mitigating this potential threat.

Missing Data Bias: Because of the large amount of data typically contained in registries, it is not feasible to meet the stringent requirements used in clinical trials. However, unlike with administrative claims data, data fields in a registry must be assessed for completeness, consistency, and accuracy to support the central activities of the registry. The NCDR Data Quality Program consists of 3 main components: data completeness, consistency, and accuracy. Completeness

focuses on the proportion of missing data within fields, whereas consistency determines the extent to which logically related fields contain values consistent with other fields. Accuracy characterizes the agreement between registry data and the contents of original charts from the hospitals submitting data. The thresholds for all critical elements in a performance measure are set high to ensure data completeness and consistency for the overall calculation of the performance measure. Therefore it is unlikely missing data bias would threaten the validity properties.

Selection Bias: In January 2005 the Centers for Mediare and Medicaid Services (CMS) expanded the covered indications for primary prevention ICDs to incorporate the findings from published literature. As part of this expansion, CMS mandated that a national registry be formed to compile data on Medicare patients implanted with primary prevention ICDs to confirm the appropriateness of ICD utilization in this patient population. CMS selected the NCDR ICD Registry as the mandated national registry in October 2005 and enrollment opened on January 1, 2006. As the CMS-mandated registry for hospitals that perform ICD implantation procedures, the ICD Registry essentially requires all hospitals that receive Medicare funding, to a participant of the NCDR Registry. This limits the potential for selection bias. Additionally, based on the entity and patient descriptive statistics, there does not appear to be certain subgroups of hospitals or patients who are excluded. Lastly, the exclusion frequencies did not appear to be unusually high.

Confounding Bias: No empirical testing was performed since this metric is neither an outcome or resource use measure.

2b3. EXCLUSIONS ANALYSIS

Note: *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*

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

2b3.1. 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*)

The only exclusions for this measure are noted under S.10. (Discharge status of expired; not eligible for either ACE/ARB or beta blockers). These exclusions are relatively rare and firmly supported by the clinical rationale.

2b3.2. 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*)

Exclusions	Patient Stays		Facilities	
Total	**665983**	**100.0**	**1709**	**100.0**
Discharge not in 2013 and 2014	420784	63.2	96	5.6
Remaining	**245199**	**36.8**	**1613**	**94.4**
Died during hospital	710	0.3	0	0.0
Remaining	**244489**	**99.7**	**1613**	**100.0**
Not eligible to the composite measure	48926	20.0	7	0.4
Study Cohort	**195563**	**80.0**	**1606**	**99.6**
The composite measure at discharge	159321	81.47	1589	98.94

2b3.3. 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*)

As noted above, there are no 'discretionary' exclusions. All exclusions are necessary to the accurate calculation of performance on the composite measure. For example, patients need to survive to discharge to be eligible for the measure. Similarly, it would be inappropriate to calculate the measure among patients ineligible for the medications.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
Note: *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used? *(check all that apply)*
☐**Endorsed (or submitted) as individual performance measures**
☐**No risk adjustment or stratification**
☐**Statistical risk model**
☐**Stratification by risk categories**
☐**Other,** Click here to enter description

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

N/A

2b4.4. What were the statistical results of the analyses used to select risk factors?

N/A

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
N/A

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
N/A

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
N/A

2b4.9. Results of Risk Stratification Analysis:
N/A

2b4.10. 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?*)
N/A

***2b4.11. 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*)
N/A

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
Note: *Applies to the composite performance measure.*
Across stratified analyses based on sex, age, race, and proportion of patients who are insured through Medicaid, we found significant overlap in the distribution of hospital performance.

2b5.1. 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)*
We examined variation in hospital performance for the composite measure based on sex, age, race, and the proportion of patients who are insured through Medicaid to identify meaningful differences.

2b5.2. 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*)

Across stratified analyses based on sex, age, race, and proportion of patients who are insured through Medicaid, we found significant overlap in the distribution of hospital performance.

2b5.3. 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?*)

Given the gaps in care, there continues to be an opportunity for improvement.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS *Note:* *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*
If only one set of specifications for each component, this section can be skipped.

Note*: This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted?*)

N/A

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH
Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

We believe the content validity of this measure has been achieved by virtue of the noted expertise of those individuals who developed this measure. The individual components of the composite have

already shown to impact clinical outcomes. However the empirical analysis demonstrating the individual component measures fit the overall quality construct is currently being researched. The testing will focus on construct validation which will test the hypothesis on the theory of the construct that following these processes for patients with ICD implantations lead to better outcomes. This research is expected to ultimately be published in the medical literature.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc.*; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

2d3. Empirical analysis demonstrating that the approach for handling missing data minimizes bias (*i.e., achieves scores that are an accurate reflection of quality*)**.**
Note: *Applies to the overall composite measure; the focus is on missing data rather than exclusions, which are considered in 2b3.*
The composite discharge medication measure is specified such that cases with missing data are assumed to have not met the metric. The performance ranges throughout this application reflect this approach. By following this method, the scores should be a true depiction of performance

scores.

2d3.1. What is the overall frequency of missing data and the distribution of missing data across providers?

As noted above, there are no "discretionary" exclusions. All exclusions are necessary to the accurate calculation of performance of the measure. See section 2b3.2.

2d3.2. Describe the method used to compare approaches for handling missing data (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

Missing data defaults to "performance not met". This measure assumes that missing documentation on the process results in a failure of meeting a evidence based therapy.

2d3.3. What were the statistical results obtained from the analysis of missing data? (e.g., *results of sensitivity analysis of effect of various rules for missing data; if no empirical analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

No empirical analysis was performed. However, it was felt that the method employed would minimize the potential for gaming.

2d3.4. What is your interpretation of the results in terms of demonstrating that the approach used for missing data minimizes bias? (i.*e., what do the results mean in terms of supporting the selected approach for missing data; if no empirical analysis, provide rationale for the selected approach for missing data*)

Given the low frequency of exclusions, we do not believe that the exclusions have any impact on the validity, accuracy or interpretability of this measure. The exclusions have little potential for bias especially given the ICD Data Quality Program audits all essentially performance measure elements on a 3 year cycle and would detect misclassifications of patient records.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

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

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
No feasibility assessment **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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Availability:
Participating hospitals report patient demographics, medical history, risk factors, hospital presentation, initial cardiac status, procedural details, medications, laboratory values and in-hospital complications. All of the data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions can manually report using a free web-based tool or automate the reporting by using certified software developed by third-party vendors. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

Sampling:
There is no sampling of patient data allowed within the contractual terms of participation in the ICD Registry in

NCDR. Section 2.b of the NCDR Master Agreement with participants includes 'Participant Responsibilities': "b. Use of ACCF Data Set and ACCF-Approved Software. Participant will submit a data record on each patient who receives medical care and who is eligible for inclusion in the Registries in which Participant is participating under this Agreement." Adult patients, ages 18 years and older, who have an ICD implanted. Patients are selected for inclusion by reviewing existing medical records and no direct interaction with the patient will be required outside of the normal course of care. There will be no discrimination or bias with respect to inclusion on the basis of sex, race, or religion.

Patient confidentiality:

Patient confidentiality is preserved as the data are in aggregate form. The ICD Registry dataset, comprised of approximately 320, data elements was created by a panel of experts using available ACC-AHA guidelines, data elements and definitions, and other evidentiary sources. Private health information (PHI), such as social security number, is collected. The intent for collection of PHI is to allow for registry interoperability and the potential for future generation of patient-level drill downs in Quality and Outcomes Reports. Registry sites can opt out of transmitting direct identifiers to the NCDR, however, so inclusion of direct identifiers in the registry is at the discretion of the registry participants themselves. When using the NCDR web-based data collection tool, direct identifiers are entered but a partition between the data collection process and the data warehouse maintains the direct identifiers separate from the analysis datasets. The minimum level of PHI transmitted to the ACCF when a participant opts out of submitting direct identifiers meets the definition of a Limited Dataset as such term is defined by the Health Insurance Portability and Accountability Act of 1996.

Data collection within the NCDR conforms to laws regarding protected health information. Patient confidentiality is of utmost concern. The proposed measure does not include a patient survey. Physician and/or institutional confidentiality are maintained by de-identified dashboard reports. There is no added procedural risk to patients through involvement in the ICD Registry. No testing, time, risk, or procedures beyond those required for routine care will be imposed. The primary risk associated with this measure is the potential for a breach of patient confidentiality. The ACCF has established a robust plan for ensuring appropriate and commercially reasonable physical, technical, and administrative safeguards are inplace to mitigate such risks.

Data are maintained on secure servers with appropriate safeguards in place. The project team periodically reviews all activities involving protected health information to ensure that such safeguards including standard operating procedures are being followed. The procedure for notifying the ACCF of any breach of confidentiality and immediate mitigation standards that need to be followed is communicated to participants. ACCF limits access to Protected Health Information, and to equipment, systems, and networks that contain, transmit, process or store Protected Health Information, to employees who need to access the PHI for purposes of performing ACCF's obligations to participants who are in a contractual relationship with the ACCF. All PHI are stored in a secure facility or secure area within ACCF's facilities which has separate physical controls to limit access, such as locks or physical tokens. The secured areas are monitored 24 hours per day, 7 days per week, either by employees or agents of ACCF by video surveillance, or by intrusion detection systems.

Each participant who has access to the NCDR website must have a unique identifier. The password protected webpages have implement inactivity time-outs. Encryption of wireless network data transmission and authentication of wireless devices containing NCDR Participant's information ACCF's network is required. Protected Health Information may only be transmitted off of ACCF's premises to approved parties, which shall mean: A subcontractor who has agreed to be bound by the terms of the Business Associate Agreement between the ACCF and the NCDR Participant.

Time of Data collection:
1 Full time employee can enter on average roughly 1200 patient records per year
(citation: ACC Marketing Intelligence Team)

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
The ACCF's program the National Cardiovascular Data Registry (NCDR) provides evidence based solutions for cardiologists and other medical professionals committed to excellence in cardiovascular care. NCDR hospital participants receive confidential benchmark reports that include access to measure macro specifications and micro specifications, the eligible patient population, exclusions, and model variables (when applicable). In addition to hospital sites, NCDR Analytic and Reporting Services provides consenting hospitals' aggregated data reports to interested federal and state regulatory agencies, multi-system provider groups, third-party payers, and other organizations that have an identified quality improvement initiative that supports NCDR-participating facilities. Lastly, the ACCF also allows for licensing of the measure specifications outside of the Registry.
It should be noted that the centers already have to participate in this specific registry for reimbursement purposes so that currently almost all hospitals that implant ICDs in Medicare populations already participate. Hence there is no additional cost.
Measures that are aggregated by ACCF and submitted to NQF are intended for public reporting and therefore there is no charge for a standard export package. However, on a case by case basis, requests for modifications to the standard export package will be available for a separate charge.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. **Current and Planned Use**

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

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with	

Benchmarking (external benchmarking to multiple organizations) Quality Improvement (Internal to the specific organization) Not in use	

4a.1. **For each CURRENT use, checked above, provide:**

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

N/A, not being publicly reported.

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

N/A, not being publicly reported.

4a.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.*)

ACC is committed to implementing this measure. ACC is an authorized organization to receive CMS data through the ResDAC application process. Unfortunately, it has been determined by ResDAC that this authorization does not permit use of CMS for performance measure reporting purposes, either to hospitals or for public display. ACC is currently in process of applying to be a Qualified Entity. It is unclear if this pathway will permit measure implementation. ACC also is commenting on and tracking proposed language in 21st Century Cures legislation, which does appear to create a pathway for use of CMS data for this type of reporting purpose.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Trends: Data presented below identify the improving trend shown by participating hospitals in prescribing each of the three medications within this composite measure. While the top 10% of performers saw slight performance improvement, the hospitals in the lower percentiles (below the median) improved significantly.

The two tables below indicate aggregated hospital performance results in 2009 (second table) and aggregated hospital performance results for this composite during Jan 2011-Dec 2012 (first table listed below)

2011 - 2012
Hospitals: n=1552
Patients: n=243186
Mean: 74%
Std Deviation: 16%
Percentiles
90th: 91%
75th (Quartile 3): 84%
50th (Median): 76%
25th (Quartile 1): 67%
10th: 56%

2013 - 2014
Hospitals: n=1606
Patients: n=195563
Mean: 78%
Std Deviation: 17%
Percentiles
90th: 97%
75th (Quartile 3): 89%
50th (Median): 79%
25th (Quartile 1): 71%
10th: 59%

Geographic area and number and percentage of accountable entities and patients included

Geographic area: This registry data captures hospital data from the United States as well as territories. The United States data are included in the aggregate. Other country data are excluded from national aggregates for the purpose of reporting.
Number of accountable entities:1552 for calendar years 2011-2012; 1606 for calendar years 2013-14.
Patients included: 243186 for calendar years 2011-2012; 195563 for calendar years 2013-2014

4b.2. **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.**
Improvements were demonstrated.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR**

has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.

Inaccuracies may occur if certified vendors export data incorrectly, in transmission of data from medical record to a paper form and then to the online data collection tool. Some sites may over-code medication exclusions.
A vendor certification process has been established to ensure high quality data collection and submission.
The NCDR Data Quality Program is in place to assess reliability of data abstraction. For additional details about the NCDR Data Quality Program please see testing supplement.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0066 : Coronary Artery Disease (CAD): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy - Diabetes or Left Ventricular Systolic Dysfunction (LVEF < 40%)
0070 : Coronary Artery Disease (CAD): Beta-Blocker Therapy-Prior Myocardial Infarction (MI) or Left Ventricular Systolic Dysfunction (LVEF <40%)
0071 : Persistence of Beta-Blocker Treatment After a Heart Attack
0081 : Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)
0083 : Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)
0117 : Beta Blockade at Discharge
0236 : Coronary Artery Bypass Graft (CABG): Preoperative Beta-Blocker in Patients with Isolated CABG Surgery
0594 : Post MI: ACE inhibitor or ARB therapy

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

Note also 0696: STS composite score. section 5.1a. has noted this has been de-endorsed, but after re-confirming with the NQF Quality Positioning System, believe this measure is still endorsed.

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
We believe the aforementioned measures are not in direct competition with measure 0965. In all cases the measure focuses on the same process, but different target population. Surgical (CABG): 0117, 0236, 0696 HF: 0083, 0081 CAD and outpatient focused: 0070, 0066 AMI: 0071 AMI, hypertension, heart failure, and diabetes: 0594 While ACC's ICD Registry does capture patient history, risk factors, and other ailments, the focus of the Registry surrounds the clinical conditions of the implantation of an ICD, dual chamber, or CRT-D device. Secondly, the Registry does not capture hypertension as an element.

5b. Competing Measures

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

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

Appendix

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

Attachment **Attachment:** icd_v2_datadictionary_codersdictionary_2-1-635699805607170839.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American College of Cardiology
Co.2 **Point of Contact:** Jensen, Chiu, comment@acc.org, 202-375-6000-
Co.3 **Measure Developer if different from Measure Steward:** American College of Cardiology
Co.4 **Point of Contact:** Jensen, Chiu, comment@acc.org, 202-375-6000-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**

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

For this particular topic those individuals who were involved in identifying the key attributes and variables for this process measure were leaders and experts in the field of electrophysiology. Serial phone calls were held to both define the eligible population and given process. These clinical leaders are noted below.

NCDR Clinical Subworkgroup ensured the measure demonstrated an opportunity for improvement, had strong clinical evidence, and was a reliable and valid measure. These members included Drs. Jeptha Curtis (Chair), Frederick Masoudi, John Rumsfeld, Matt Reynolds, and Mark Kremers.

NCDR Scientific Quality and Oversight Committee—a committee that served as the primary resource for crosscutting scientific and quality of care methodological issues. These members included Drs. Frederick Masoudi (Chair) , David Malenka, Thomas Tsai, Matthew Reynolds, David Shahian, John Windle, Fred Resnic, John Moore, Deepak Bhatt, James Tcheng, Jeptha Curtis, Paul Chan, Matthew Roe, and John Rumsfeld.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2011

Ad.3 **Month and Year of most recent revision:** 02, 2015

Ad.4 **What is your frequency for review/update of this measure?** With dataset revisions and based on new evidence.

Ad.5 **When is the next scheduled review/update for this measure?** 2016

Ad.6 **Copyright statement:** American College of Cardiology Foundation All Rights Reserved

Ad.7 **Disclaimers:** ACC realizes the various NCDR endorsed measures are not readily available on their own main webpage. However, ACCF plans to update their main webpage (acc.org) to include the macrospecifications of the NQF endorsed measures. ACC hopes to work collaboratively with NQF to create a consistent and standard format would be helpful for various end users. In the interim, the supplemental materials include the details needed to understand this model. In addition, interested parties are always able to contact comment@acc.org to reach individuals at the ACC Quality Measurement Team.

Ad.8 **Additional Information/Comments:** ACC appreciates the opportunity to submit measures for this NQF endorsement maintenance project.

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 #: 2396
De.2. **Measure Title:** Carotid artery stenting: Evaluation of Vital Status and NIH Stroke Scale at Follow Up
Co.1.1. **Measure Steward:** American College of Cardiology
De.3. **Brief Description of Measure:** Proportion of patients with carotid artery stenting procedures who had follow up performed for evaluation of Vital Status and neurological assessment with an NIH Stroke Scale (by an examiner who is certified by the American Stroke Association) Occurring between day 21 and the end of day 60 after the procedure. (Days 21-60 inclusive)
1b.1. **Developer Rationale:** This measure is important to determine the number of patients that are being followed after a carotid artery stent procedure. Specifically vital status and if a NIHSS were completed at follow-up.

S.4. **Numerator Statement:** Patient Status (alive or Deceased) at follow-up AND Neurologic status with an assessment using the NIH Stroke Scale (by an examiner who is certified by the American Stroke Association) AND Discharge Status (alive or Deceased)
S.7. **Denominator Statement:** Count of CARE Registry patients that had a carotid artery stenting procedure
S.10. **Denominator Exclusions:** Patients with a discharge status of deceased
Patients with was an acute, evolving stroke and dissection during the episode of care

De.1. **Measure Type:** Process
S.23. **Data Source:** Electronic Clinical Data : Registry
S.26. **Level of Analysis:** Facility, Population : National

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

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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.

The developer provides the following evidence for this process measure:

- This is facility- and population-level measure calculates the proportion of patients with carotid artery stenting (CAS) procedures who had follow up performed for evaluation of Vital Status and neurological assessment with an NIH Stroke Scale (NIHSS) between 21 and 60 days inclusive after the procedure by an examiner who is certified by the American Stroke Association.
- The evidence finds higher recurrent stenosis after carotid artery stenting than carotid endartectomy requiring stroke symptom assessment with a reliable and standardized tool, NIHSS. They also state mortality and stroke are common within 2 years post-CAS, and more common immediately following CAS. The developer provides a

consensus recommendation – categorized as a guideline in the submission – with 7 articles for attaining facility & operator procedure competence (no grading is assigned). This document recommends post-procedural follow-up and monitoring using standardized tools and definitions, with a neurologic assessment performed by a qualified and NIH Stroke Scale-certified individual for all patients undergoing carotid stenting. The developer also provides a clinical guideline (Class IIa) that recommends *noninvasive imaging* 1 month, 6 months, and annually after carotid revascularization. Grading is based on the 2010 Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guideline. Thirty-day post-CAS neurological raassessment with NIHSS certified examiner is recommended.

- The developer provides the Quantity/Quality/Consistency for the imaging clinical guideline, and they also include a 2013 article outlining comprehensive stroke center quality metrics, though this measure does not appear in the article.
- A process to outcomes logic diagram states reassessment 30-days after CAS with NIHSS assists in determining the disability caused by the stroke and guides patient-specific treatment planning. The evidence states disability reassessment with NIHSS should 30 days post-CAS, while the specifications state "between 21 and 60 days" inclusive.

Questions for the Committee:

- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- The developer provides the following performance table of NCDR CARE Registry from 2007Q1 through 2013 Q1 demonstrating gaps in care:

	Total	Both Vital and NIHSS at F/U		
	n = 18212	1 n = 7220	0 n = 10992	P-Value
Followup Measures				
Has_vital	12883 (70.74%)	7220 (100.00%)	5663 (51.52%)	< 0.001
has_nihss	7220 (39.64%)	7220 (100.00%)	0 (0.00%)	< 0.001

- It is not clear if the "has nihss" includes patients who were deceased at the time of the follow-up (21 to 60 days inclusive post-CAS).
- The developer also provides post-CAS morbidity and mortality data, as well as stroke and stroke mortality data.
- Disparities data is also provided by age, sex, and race, as well as multiple clinical findings (e.g., comorbidities, testing and medications). The developer also provides data on hospital type and rural/suburban hospital characteristics. Insurance status & type are captured on the measure data collection tool.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - The developer provides a consensus recommendation – categorized as a guideline – with 7 articles for attaining facility and operator procedure competence (no grading is assigned). It appears to be in the category of expert opinion Level of evident: C primary source of recommendation was consensus opinion, case studies or standard of care
   - The measure looks at the important point of evaluating Vital status and NIHSS stroke evaluation post carotid artery stenting.
   - The cited evidence does not appear to support the measure focus.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***
   - Clinical Practice Guideline Recommendation Monitoring 1a of outcomes with independent post-procedural neurological assessment using standardized instruments and definitions is critically important to ensure high-quality intervention and patient safety. Institutions offering carotid stent placement must have a quality assurance program specifically designed to assess the results of carotid interventions in their locale. The integrity and accuracy of outcome reporting is reliant on the incorporation of mandatory independent and objective neurologic assessment by a qualified and NIH Stroke
   - Facility and population level measure looking at CAS patients who had a Vital status evaluation and NIHSS evalution at 21-60 days.
     - The restenosis rate after CAS as compared to CEA doesn't necessarily correlate with stroke rate.
     - Some of the articles presented for support the measure are dated.
   - The evidence cites a higher rate of recurrent carotid artery stenosis >70% after carotid stenting than endarterectomy during a 2-year follow-up period. This is not directly applicable to this process measure, which assesses rates of 21-to-60-day follow-up with a neurological assessment (standardized NIHSS stroke scale) performed by an examiner certified by the American Stroke Association (ASA) and ascertainment of mortality status.
   - The process of care does not seem proximal, nor closely related to desired outcomes (lower rates of recurrent stenosis and stroke).
   - I am not aware of any logical basis for possible exception to the evidence criterion, with respect to performance measures of a related health outcome or evidence-based intermediate clinical outcomes, intervention, or treatment. The developers cite general support from the ASA and Society of Vascular and Interventional Neurology for quality of care metrics for comprehensive stroke centers as relevant expert opinion.
   - I consider it to be generally unacceptable to hold providers accountable for measure performance without empirical evidence of benefit from improving performance on the measure.

1b. ***Committee's Comments on Performance Gap:***
   - Overall considerable variation or less than optimal performance in the quality of care and disparities in care across population groups.
   - The performance gap would be that all patients undergoing CAS have Vital and NIHSS evaluation. The perfomance table provided is not clear and does not delineate if deceased patients were included in followup.
   - The developers cite performance rates of 0-100%, which would appear to warrant a national performance measure, if other measure criteria were met.
   - I did not see in the application and am not aware of evidence for significant disparities in care.
   - I do not think it should be indicated as disparities sensitive.

1c. ***Committee's Comments on Composite Performance Measure:***
   - Not Applicable
   - Not clearly stated.
   - The numerator elements of this measure (vitality status and neurological assessment) are logically constructed performance measures, with a reasonably well articulated rationale. However this does not appear to be a true

composite measure, since the only element within clinician control in the timeframe is the neurological assessment.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- This is a hospital/acute care facility level measure that requires documentation of a completed NIHSS assessment between 21 and 60 days inclusive post-CAS procedure by an American Stroke Association certified examiner.
- The developer is encouraged to clarify the denominator population as patients age is not provided in the specifications, though the narrative in section 3c.1 states adult patients, 18 years and older.
- The denominator exclusions include patients with a discharge status of deceased, and patients with an acute, evolving stroke and dissection during the episode of care. It is not clear if the episode of care is at any time during the hospitalization when the CAS was performed, or at any time after the CAS was performed.
- The developer should clarify if documentation of deceased vital status prior to the 21 to 60 days inclusive follow-up is considered as a performed quality action (equals a performance pass).
- The evidence and collection tool provided states the NIHSS should be conducted by someone other than the operator for the current procedure, and by an American Stroke Association certified examiner. It is not clear how these characteristics are assessed.
- The data collection tool appears to allows patient reasons (e.g., patient refusal, patient unavailable and other) for no follow-up, which does not appear in the specifications.
- A brief calculation algorithm is provided.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate clinical concepts included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- The developer provides 2 types of reliability testing including the signal-to-noise facility-level testing of the measure score for facilities who completed neurological function testing, and a test-retest methodology to test data element reliability of patient characteristics only.
  The signal-to-noise analysis, which is appropriate for this type of measure, differentiates the true difference between measured entities (the signal) to random measurement error (the noise). A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in between hospital performance. A value of 0.7 is often regarded as a minimum acceptable reliability value. The table demonstrates very high reliability and is provided based on CAS procedural volumes, with reliability increasing in higher volumes.

Level	Signal-to-Noise
All, >10 Procedures	.984
>Q1 (>55 Procedures)	.986

>Q2 (>91 Procedures)	.989
>Q3 (>161 Procedures)	.993
>Average (>131 Procedures)	.992

- The developers also used a test-retest of assess the accuracy of the data collection methodology and applicable patient characteristic data elements entered into the CARE registry including age, gender, race, smoking, history of PAD, diabetes, chronic lung disease & dyslipidemia. To construct the test/retest, developers identified 449 patient during the testing period with 2 procedures in facilities completing >30 procedures to produce more reliable estimates of the data elements. Data elements included in this testing that are not applicable to measure calculation, are used to testing accuracy and reproducibility of data collection.
- Data element definitions and differences are provided for each data element. Age, gender and race did not vary between the test/retest populations, and smoking, history of PAD, diabetes, chronic lung disease & dyslipidemia variation were < 2.7%.
- Patients with missing data were not included in either of the tests.

Questions for the Committee:

- *Does limiting testing to > 30 procedures per facility impact the measure reliability?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The measure specifications allow for follow-up and NIHSS reassessment between 21-60 days inclusive, and the evidence states 30 days.
- NIHSS completed by a certified examiner who is not the CAS operator is not included in the measure.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- The developer provided content validity stating numerous studies assess the comparative effectiveness of stenting vs. surgery for carotid occlusion on the outcomes of myocardial infarction, death and stroke.
- The developer provided face validity by expert cardiologist panel who actively perform CAS and carotid endartectomy procedures including leading experts in the field and vetted the measure with 3 committees and the 16 member NCDR Management Board and 31 member ACCF Board of Trustees prior to NQF submission.
- No formal statistical validity tests were provided.

Questions for the Committee:

- *Do the face and content validity provided demonstrate the measure has sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The exclusions include patients deceased at discharge and patients with an acute, evolving stroke and

dissection.

- The developer does not provide statistical testing methods and results from excluding patients with an acute stroke. Instead the developer provides a table that shows the differences in clinical characteristics of those with an acute evolving stroke as compared with those treated for stable carotid disease.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This measure is not risk adjusted.

2b5. Meaningful difference:

- The developer notes an association between procedural volume and follow up rates; higher volume centers perform substantially greater than poor performing centers. The developer also states that performance rates range from 0% to 100% therefore it is feasible for hospitals to assess all of their patients and NQF endorsement of this measure would increase hospitals' commitment.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- This is not applicable as there is only one data source/specification for this measure.

2b7. Missing Data

- The developer describes the NCDR Data Quality Report that assesses for data completeness (missing data), and consistency and accuracy (integrity), scoring provided data with 3 "light" levels: Red lights (submission failure for integrity and completeness check – data not processed), yellow lights (passed for integrity and fails completeness – requires data resubmission), and green lights (passed all quality checks and included for aggregate computations). No sampling of NCDR patient data is allowed as registry inclusion mandates 100% consecutive patients.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- Measure is well defined and specified. Should produce consistent, reliable results about quality of care. Reliability testing for the elements assessed appeared consistent and accurate in the Care Registery.
- Data elements are not clearly defined.
- The measure data elements lack clarity for the denominator (age, method of ascertaining ASA examiner certification and non-procedure-operator status of the conductor of the NIHSS tool, and credit status for documentation of death prior to 21 days are missing).
- The algorithm is clear, but seems oversimplified.
- It seems unlikely that the measure can be consistently implemented.

2a2.: ***Committee's Comments on Reliability-Testing:***

- Big variation s in collection of 30 day outcomes in hospitals from 0 to 100%.
- The patients included did test with an overall .992 reliability on a sigal to noise analysis. However, patients with missing data were not included in the testing.
- Limiting testing to facilities performing >30 procedures probably inflates the measure reliability.
- I am unsure whether the test sample of 449 patients is sufficient for widespread implementation.
- The results appear to be sufficiently reliable to identify differences in performance.

2b1.: ***Committee's Comments on Validity-Specifications***:

- Measure was shared with a noted panel of expert cardiologists wh participate regularly in carotid artery stenting

procedures. All members of the committee reported that the measures appear to be a good indication of positive outcome in CAS procedures.
- Specifications are not consistent with the evidence.
- The specifications seem inconsistent with the evidence for the non-CAS operator being the certified examiner. I think the 21-60 day timeframe is a reasonable window around the 30-day expert opinion point. (The evidence actually comes from a 2-year timeframe for inferior stenting restenosis outcome.)

2b2.: ***Committee's Comments on Validity-Testing:***
- There was content validity from prior studies and expert consensus bit no statistical validity was provided.
- The face and content validity arguments are based on expert opinion only and do not appear to support validity of the measure as a quality indicator.

2b3-7.: ***Committee's Comments on Threats to Validity:***
- The exclusions are significant and patients or patient groups may be excluded form the the measure.
- The measure is not risk adjusted
- The exclusions make sense, but there is no analysis of the effects of the exclusions. It is unclear whether the exclusions/exceptions are frequent enough and/or vary enough across providers to be needed and worth the data collection burden.
- It is unclear whether this measure identifies meaningful differences in quality of post-CAS care.
- This measure is not risk adjusted.
- The measure would appear to address meaningful differences in quality, based on their reporting a performance range of 0-100%.
- Ironically, I found no data on the potential effect of missing data.
- This was not presented as a composite measure.

2d.: ***Committee's Comments on Composite Performance Measure:***
- Not Applicable

Criterion 3. Feasibility

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

- The data source used to collect and calculate measure performance is the NCDR Care Registry. The developer states that ALL data elements are in defined fields in electronic clinical data, and may be collected via third-party vendors. The specifications are available in the public domain.
- Though it is not apparent, it is assumed the post-discharge data is captured by the facility reporting the measure.
- The developer provides the data collection strategy including costs and fees noting for calendar year 2014 the annual pricing for hospitals, NCDR Analytic and Reporting Services, and licensing of measure specifications ranges from $2900-$50,000.

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***
- By Product of care
- The data can be obtained from the NCDR CARE registry. In using the registry for projects in the past, there is a likelihood that not all the data was captures.
- It is unclear whether the ASA examiner certification status of the conductor of the NIHSS tool is routinely generated and used during care delivery. The developers did not specify how this data entry and verification of

this key element would be handled.
- Most, but probably not all required data elements are available in electronic form.
- It appears to be dubious whether the data collection strategy is fully ready to be put into operational use.

Criterion 4: Usability and Use

4. Usability and 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.

- It is unclear if the measure is currently or planned for use with ACC's CARE Registry of the National Cardiovascular Data Registry (NCDR).
- The measure is not currently publicly reported but the developers state that the plan is for this measure to be publicly reported in the future. The developer provides a plan for future use/ implementation and are applying to be a CMS Qualified Entity to allow developers access to CMS reporting. Further clarification is needed.
- The measure may be used for quality improvement purposes.
- The developer states there are no unintended consequence identified for the measure.

***Questions for the Committee**:*
- *Is the measure publicly reported?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***
- Public Reporting is planned. Current use is quality improvement.
- Not currently publicly reported, but plans for being a CMS Qualified entity in the future to allow for reporting is possible. The actual utilizaiton of this measure should be further clarified.
- The measure is not currently publicly reported, and it is unclear whether it will be used by the NCDR.
- The developers say that the ACC is committed to implementing this measure, but they also say that CMS data cannot be used for performance measure reporting to hospitals or the public.
- It is unclear how the performance results would be used to further the goal of high quality efficient healthcare.
- The benefits of the measure are unclear.
- The developers identified no unintended consequences, and I can think of none, but since unintended consequences are often unanticipated, that does not say much.

Criterion 5: Related and Competing Measures

- The developer provided no related or competing measures.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Title Carotid artery stenting: Evaluation of Vital Status and NIH Stroke Scale at Follow Up

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: 12/18/ 2013

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

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

Subcriterion 1a. Evidence to Support the Measure Focus

The measure focus is a health outcome or is evidence-based, demonstrated as follows:

- Health outcome:[3] a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome, Process,[4] or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome.
- Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes.
- Efficiency:[6] evidence for the quality component as noted above.

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. 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.
5. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
6. Measures of efficiency combine the concepts of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of:

Outcome

☐ Health outcome: Click here to name the health outcome

Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)

☐ Intermediate clinical outcome: Click here to name the intermediate outcome

☒ Process: 30 Day Follow-up Assessment of Stroke/Death after Carotid Artery Stent Revascularization

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

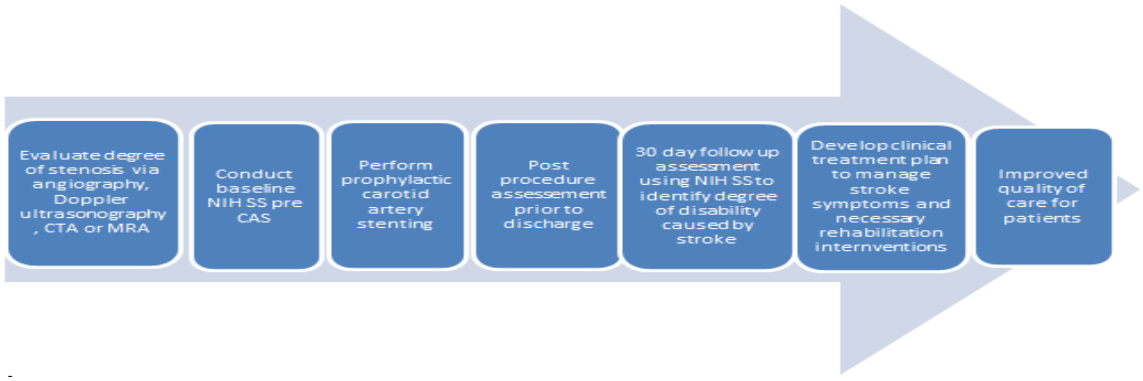
1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (including date) and **URL for guideline** (if available online):

#1) Rosenfield K, Cowley MJ, Jaff MR, Ouriel K, Gray W, Cates CU, Feldman T, Babb JD, Gallagher A, Green R, Kent KC, Roubin GS, Weiner BH, White CW. SCAI/SVMB/SVS clinical competence statement on carotid stenting: training and credentialing for carotid stenting— multispecialty consensus recommendations, a report of the SCAI/SVMB/SVS writing committee to develop a clinical competence statement on carotid interventions. J Am Coll Cardiol 2005;45:165–74.

URL for Clinical Competence Statement: http://content.onlinejacc.org/article.aspx?articleid=1136222

#2) Brott TG, Halperin JL, Abbara S, et al. 2011 ASA/ACCF/AHA/AANN/AANS/ACR/ASNR/CNS/SAIP/SCAI/SIR/SNIS/SVM/SVS Guideline on the Management of Patients With Extracranial Carotid and Vertebral Artery Disease: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, and the American Stroke Association, American Association of Neuroscience Nurses, American Association of Neurological Surgeons, American College of Radiology, American Society of Neuroradiology, Congress of Neurological Surgeons, Society of Atherosclerosis Imaging and Prevention, Society for Cardiovascular Angiography and Interventions, Society of Interventional Radiology, Society of NeuroInterventional Surgery, Society for Vascular Medicine, and Society for Vascular Surgery Developed in Collaboration With the American Academy of Neurology and Society of Cardiovascular Computed Tomography. J Am Coll Cardiol. 2011;57(8):e16-e94.URL for guideline: http://content.onlinejacc.org/article.aspx?articleid=1144187

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

#1) Page 9 of 10 “Monitoring of outcomes with independent post-procedural neurological assessment using standardized instruments and definitions is critically important to ensure high-quality intervention and patient safety. Institutions offering carotid stent placement must have a quality assurance program specifically designed to assess the results of carotid interventions in their locale. The integrity and accuracy of outcome reporting is reliant on the incorporation of mandatory independent and objective neurologic assessment by a qualified and NIH Stroke Scale-certified individual for all patients undergoing carotid stenting”.

#2) Page 40 of 79, e55 “Noninvasive imaging of the extracranial carotid arteries is reasonable 1 month, 6 months, and annually after revascularization to assess patency and exclude the development of new or contralateral lesions. Once stability has been established over an extended period, surveillance at extended intervals may be appropriate. Termination of surveillance is reasonable when the patient is no longer a candidate for intervention”.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

#1) No grade assigned

#2) Class IIa,(Definition of Class IIa: Weight of evidence/opinion is in favor of usefulness/ efficacy. IT IS REASONABLE to perform procedure/administer treatment.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

#1) No grades in Competence Statement

#2) See table below.

ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT	CLASS I *Benefit >>> Risk* Procedure/Treatment **SHOULD** be performed/ administered	CLASS IIa *Benefit >> Risk* *Additional studies with focused objectives needed* **IT IS REASONABLE** to perform procedure/administer treatment	CLASS IIb *Benefit ≥ Risk* *Additional studies with broad objectives needed; additional registry data would be helpful* Procedure/Treatment **MAY BE CONSIDERED**	CLASS III *No Benefit* or CLASS III *Harm* Procedure/ Test / Treatment COR III: No benefit / Not Helpful / No Proven Benefit COR III: Harm / Excess Cost w/o Benefit or Harmful / 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
Suggested phrases for writing recommendations	should is recommended is indicated is useful/effective/beneficial	is reasonable can be useful/effective/beneficial is probably recommended or indicated	may/might be considered may/might be reasonable usefulness/effectiveness is unknown/unclear/uncertain or not well established	COR III: No Benefit is not recommended is not indicated should not be performed/ administered/ other is not useful/ beneficial/ effective
Comparative effectiveness phrases†	treatment/strategy A is recommended/indicated in preference to treatment B treatment A should be chosen over treatment B	treatment/strategy A is probably recommended/indicated in preference to treatment B it is reasonable to choose treatment A over treatment B		COR III: Harm potentially harmful causes harm associated with excess morbidity/mortality should not be performed/ administered/ other

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

For Recommendation #2) ACCF/AHA Task Force on Practice Guidelines. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines. American College of Cardiology Foundation and American Heart Association, Inc. Cardiosource.com. 2010. Available at: http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf and http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

The evidence review focused on the importance of 30 day follow up assessment for this patient population to determine morbidity in terms of incidence and prevalence of stroke and mortality that may be associated with carotid artery revascularization via stenting.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Level of Evidence: C (Definition of Evidence Level C: primary source of the recommendation was consensus opinion, case studies, or standard of care).

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Guideline #2) listed in section 1.a. 4.2

Evidence Level A: data were derived from multiple randomized clinical trials or meta-analyses.

Evidence Level B: data were derived from a single randomized trial or nonrandomized studies.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 2001 - 2006

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

One multinational, prospective, randomised study of 1,214 patients.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Certainty or confidence in the estimates (verbatim from Eckstein et al.,cited below)

In both the intention-to-treat and per-protocol analyses the Kaplan-Meier estimates of ipsilateral ischaemic strokes up to 2 years after the procedure and any periprocedural stroke or death do not differ between the carotid artery stenting and the carotid endarterectomy groups (intention to treat 9.5%vs 8.8%; hazard ratio (HR) 1.10, 95%CI 0.75 to 1.61; log-rank p=0.62; per protocol 9.4%vs 7.8%; HR 1.23, 95%CI 0.82 to 1.83; log-rank p=0.31).

In both the intention-to-treat and per-protocol populations, recurrent stenosis of 70% or more is significantly more frequent in the carotid artery stenting group compared with the carotid endarterectomy group, with a life-table estimate of 10.7% versus 4.6% (p=0.0009) and 11.1% versus 4.6% (p=0.0007), respectively.

Eckstein H.H., Ringleb P., Allenberg J.R.; Results of the Stent-Protected Angioplasty versus Carotid Endarterectomy (SPACE) study to treat symptomatic stenoses at 2 years: a multinational, prospective, randomised trial. *Lancet Neurol*. 7 2008:893-902.

Indirectness of studies to the measure focus

The guidelines cited in 1a.4.1. stress the importance of follow up to be conducted upon this patient population. While indirectly implied, it can be considered a natural aspect of this follow up process to determine the mortality status of a patient for this follow up visit. In addition to mortality status, this measure requires the NIH Stroke Scale to be performed during the follow up visit. This can also be an implied aspect of the follow up exam assessing for morbidity.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Only two incidences of recurrent stenoses after carotid artery stenting led to neurological symptom. After 2 years' follow-up, the rate of recurrent ipsilateral ischaemic strokes reported in the SPACE trial is similar for both treatment groups

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

One outcome identified by this study is that the incidence of recurrent carotid stenosis at 2 years (identified by ultrasound), was significantly higher after carotid artery stenting then when a CEA was performed. This reinforces the importance of continued follow up on this patient population.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.

In a Statement for Healthcare Professionals From the American Heart Association/American Stroke Association and endorsed by the Society of Vascular and Interventional Neurology, several metrics were proposed intended to provide a framework for standardized data collection at comprehensive stroke centers (CSCs) to facilitate local quality improvement efforts and to allow for analysis of pooled data from different CSCs that may lead to development of national performance standards for CSCs in the future.

1a.8.1 What process was used to identify the evidence?

A literature search was performed of the medical database www.UpToDate.com and the American Heart Association/American Stroke Associate Stroke [http://stroke.ahajournals.org] webpage using keywords: "Carotid Artery Stenosis" and "Follow Up".

1a.8.2. Provide the citation and summary for each piece of evidence.

The following is quoted verbatim from the reference cited.

Reference: Leifer D, Bravata DM, Connors JJ 3rd, Hinchey JA, Jauch EC, Johnston SC, Latchaw R, Likosky W, Ogilvy C, Qureshi AI, Summers D, Sung GY, Williams LS, Zorowitz R; on behalf of the American Heart Association Special Writing Group of the Stroke Council, Atherosclerotic Peripheral Vascular Disease Working Group, Council on Cardiovascular Surgery and Anesthesia, and Council on Cardiovascular Nursing. Metrics for measuring quality of care in comprehensive stroke centers: detailed follow-up to Brain Attack Coalition comprehensive stroke center recommendations: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2011;42: Retrieved from http://stroke.ahajournals.org/content/early/2011/01/13/STR.0b013e318208eb99.full.pdf+html on December 4, 2013

Metric 10

Percentage of patients undergoing carotid endarterectomy (CEA), or carotid angioplasty or stenting, with stroke or death within 30 days of the procedure.

Numerator: Number of patients who have a stroke or die within 30 days of CEA, or who have carotid angioplasty or stenting performed because of atherosclerotic disease.

Denominator: Total number of patients who undergo CEA or who undergo carotid angioplasty or stenting because of atherosclerotic disease.

The metric should be calculated for all procedures taken together and separately for the following groups of patients:

(1) symptomatic patients undergoing endarterectomy;

(2) symptomatic patients undergoing carotid angioplasty or stenting;

(3) asymptomatic patients undergoing endarterectomy; and

(4) asymptomatic patients undergoing carotid angioplasty or stenting.

Strokes should be included if they meet the clinical definition of a focal neurological deficit that persists for > or equal to 24 hours without other cause or if there is a focal deficit that lasts for a shorter period of time but is associated with an appropriately located acute ischemic lesion on MRI. Clinically silent acute lesions detected on diffusion-weighted MRI should not be included as complications, because they are likely to be common when MRI is performed, although their incidence and clinical significance are uncertain. Patients with confusion or encephalopathy who have multiple punctate lesions that together may explain their clinical findings should also be included as having had a stroke. Published clinical trials about complications after carotid procedures and other interventions have typically used clinical stroke as the end point and other ongoing trials also are using clinical end points. This definition of stroke will apply to this metric and subsequent ones.

This metric is limited to patients with atherosclerotic disease to ensure that the metric encompasses a uniform population of patients.

Justification: The AHA/ASA guidelines for patients with recent TIA or ischemic stroke within the past 6 months and ipsilateral severe (70% to 99%) carotid artery stenosis recommend endarterectomy by a surgeon with a perioperative morbidity and mortality rate of <6% ***(Class I; Level of Evidence A)***. For patients with recent TIA or ischemic stroke and ipsilateral moderate (50% to 69%) carotid stenosis, CEA is recommended, depending on patient-specific factors such as age, sex, comorbidities, and severity of initial symptoms if the perioperative morbidity and mortality risk is estimated to be <6% ***(Class I; Level of Evidence B)***.

Among patients with symptomatic severe stenosis (>70%) in whom either the stenosis is difficult to access surgically, medical conditions are present that greatly increase the risk for surgery, or other specific circumstances exist such as radiation-induced stenosis or restenosis after CEA, the use of carotid angioplasty and stent placement is not inferior to endarterectomy and may be considered ***(Class IIb; Level of Evidence B)***. The procedure is reasonable when performed by operators with established periprocedural morbidity and mortality rates of 4% to 6% ***(Class IIa; Level of Evidence B)***.

The role of carotid angioplasty/stenting in asymptomatic patients has not been established. The AHA/ASA "Guidelines for the Primary Prevention of Stroke" state, "The usefulness of CAS [carotid angioplasty/stenting] as an alternative to CEA in asymptomatic patients at high risk for the surgical procedure is uncertain ***(Class IIb;***

Level of Evidence C)." In this setting, if centers choose to perform carotid angioplasty/stenting on asymptomatic patients, the 30-day rate of stroke and death should be tracked separately for such patients and monitored carefully.

The recommended end point to be ascertained after carotid angioplasty and stent placement is any stroke or death within 30 days, to remain consistent with the data collected for CEA. This end point has been used in trials of carotid angioplasty and stenting. For comparable patients, the complication rate for stenting should be similar to that for endarterectomy if stenting is to be a reasonable option. In particular, the complication rate should be expected to be between 4% and 6% for symptomatic >70% stenosis. If carotid angioplasty and stenting are performed, therefore, careful attention must be paid to complication rates, so it is important for CSCs to monitor these rates.

The risk of stroke and death after carotid revascularization are important and can substantially influence the net benefit of the procedure. Assessment and reporting of the "outcome" of stroke for carotid revascularization procedures is not consistent in the absence of a clinical assessment using a standardized stroke scale, or by using claims data. A class IIa, LOE: C guideline advises noninvasive imaging of the extracranial carotid arteries be performed at 1 month, 6 months, and annually after revascularization to assess patency and exclude the development of new or contralateral lesions, it can be implied that patients will have a clinic/office follow-up visits as a follow-up to revascularization procedures. This office visit provides the opportunity for appropriate clinical assessment for key revascularization endpoints, including stroke or death. A process measure that uses a standard assessment of neurologic function, by an examiner who is certified by the American Stroke Association, is a measure that provides feedback on the ability to clearly and accurately assess for, capture and report the incidence of stroke after carotid revascularization procedures. The NIHSS is both reliable and valid, and has become a standard stroke impairment scale for use in both clinical trials and as part of clinical care in the United States [6].

When centers that perform carotid revascularization properly assess patients for adverse events (particularly for stroke) after carotid revascularization, they trigger further evaluation, if necessary. If the 30 day NIH stroke scale is (1) changed from baseline; or (2) abnormal in absence of a baseline, pre-procedure exam, then there should be some documentation on whether or not the abnormal stroke scale represents a new clinical neurological event, and should result in an evaluation by a neurologist.

According to the CARE Registry institutional outcomes reports, the median length of stay for CAS and CEA procedures is one day. This short hospital stay reflects difficulty in reporting "in-hospital" stroke outcomes as a relevant measure. Following carotid artery stenting, patients are typically discharged in one to two days. In a study that evaluated the timing of complications following CAS, 53% of postoperative events/complications occurred within 6 hours of CAS, 5.3% between 6 and 12 hours, 8% between 12 and 24 hours, and 34.2% >24 hours post procedure [1]. Late events >24 hours were access-site-related and neurologic events.

The primary endpoints of major contemporary trials used 30 day events (stroke, MI* or death) and included neurologic evaluation to identify stroke. Based on trial endpoints, 30 day outcomes have greater importance. Post-procedure stroke is one of the major adverse outcomes from carotid artery stenting and carotid endarterectomy. For example, this was the major outcome in the recent CREST trial, a randomized comparison of carotid artery stenting and surgical endarterectomy. A recent meta-analysis by Murad et al. summarized the results of 13 randomized controlled trials to assess the comparative effectiveness of stenting vs. surgery for carotid occlusion on the outcomes of myocardial infarction, death and stroke [13]. The latter of these

outcomes are proposed to serve as a process measure for quantifying the quality of carotid revascularization by this measure.

There is a sound clinical rationale for systematically measuring the outcomes of carotid revascularization. First, without knowing the outcomes, a hospital cannot know if it is applying its treatment in a safe and effective manner. Given how infrequently current providers assess the 30-day survival and stroke outcomes, it is obvious that more than half of these hospitals have no foundation with which to assess the quality of their care. We have proposed a process measure, merely assessing the stroke-free survival of treated patients, because without more clear ascertainment of outcomes it is not possible to provide risk-adjusted comparisons across centers and provide clear benchmarks of performance to identify hospitals that have the opportunity to improve. Second, as the country seeks to support the use of evidence-based medicine, the majority of the evidence in carotid disease comes from clinical trials. However, many of the trials establishing the benefits of carotid revascularization require that centers document a certain success rate, without complications of stroke or death, before the center can participate in a clinical trial. If a center does not know its rate, it will not know whether or not the benefits observed in a clinical trial apply to their practice. Finally, for patients to be adequately informed about the risks and benefits of treatment, hospitals need to have reliable data to share with their patients. By collecting, analyzing and reporting the outcomes of treatment, hospitals will be much better able to provide their patients the information that they need to make a treatment decision.

Stroke is the second leading cause of all hospital admissions among older patients and the leading reason for neurology-related admissions. From 1999 to 2009, the number of inpatient discharges from short stay hospitals with stroke as the first-listed diagnosis has remained stable with 961,000 discharges in 1999 and 971,000 discharges in 2009 (National Heart, Lung, and Blood Institute [NHLBI] tabulation, National Hospital Discharge Survey [NHDS], National Center for Health Statistics [NCHS]). Correspondingly, stroke death rates fell by 24% from 1994 to 2004. This decline suggests that there have been general improvements in the management of patients with acute stroke, decreases in the severity of stroke and/or improved detection or coding of milder stroke cases. Part of the decline in hospital stroke mortality may be due to the shorter length of stay resulting in more out of hospital death. The greatest risk of mortality for patients with stroke occurs in the first 30 days, with case-fatality rates ranging from 8% to 20% for ischemic stroke, with substantially higher rates for stroke due to subarachnoid or intracerebral hemorrhage (as high as 50%). The immediate cause of death in more than 60% of stroke cases is thought to be related to complications of the stroke itself. After the first week, cardiac causes, pneumonia, pulmonary embolism, sepsis, and other medical complications account for the majority of the stroke-related mortality. In 2008, approximately 46% of all stroke deaths occurred in the hospital (unpublished NHLBI tabulation of NCHS 2008 Mortality Data Set). The annual U.S. economic burden of stroke is estimated at $20.4 billion for direct and $53.6 billion indirect costs. [7]

1] Timing and frequency of complications after carotid artery stenting: what is the optimal period of observation? Tan KT, Cleveland TJ, Berczi V, McKevitt FM, Venables GS, Gaines PA. J Vasc Surg. 2003;38(2):236

2] 30 day results from the SPACE trial of stent-protected angioplasty versus carotid endarterectomy in symptomatic patients: a randomised non-inferiority trial. SPACE Collaborative Group, Ringleb PA, Allenberg J, Brückmann H, Eckstein HH, Fraedrich G, Hartmann M, Hennerici M, Jansen O, Klein G, Kunze A, Marx P, Niederkorn K, Schmiedt W, Solymosi L, Stingele R, Zeumer H, Hacke W. Lancet. 2006;368(9543):1239

3] Endarterectomy versus stenting in patients with symptomatic severe carotid stenosis. Mas JL, Chatellier G, Beyssen B, Branchereau A, Moulin T, Becquemin JP, Larrue V, Lièvre M, Leys D, Bonneville JF, Watelet J, Pruvo JP, Albucher JF, Viguier A, Piquet P, Garnier P, Viader F, TouzéE, Giroud M, Hosseini H, Pillet JC, Favrole P, Neau JP, Ducrocq X, EVA-3S Investigators. N Engl J Med. 2006;355(16):1660.

4] Carotid artery stenting compared with endarterectomy in patients with symptomatic carotid stenosis (International Carotid Stenting Study): an interim analysis of a randomised controlled trial. International Carotid Stenting Study investigators, Ederle J, Dobson J, Featherstone RL, Bonati LH, van der Worp HB, de Borst GJ, Lo TH, Gaines P, Dorman PJ, Macdonald S, Lyrer PA, Hendriks JM, McCollum C, Nederkoorn PJ, Brown MM. Lancet. 2010;375(9719):985.

5] Thirty-day outcomes for carotid artery stenting in 6320 patients from 2 prospective, multicenter, high-surgical-risk registries. Gray WA, Chaturvedi S, Verta P, Investigators and the Executive Committees. Circ Cardiovasc Interv. 2009;2(3):159.

6] Goldstein LB, Bushnell CD, Adams RJ, Appel LJ, Braun LT, Chaturvedi S, Creager MA, Culebras A, Eckel RH, Hart RG, Hinchey JA, Howard VJ, Jauch EC, Levine SR, Meschia JF, Moore WS, Nixon JV, Pearson TA; on behalf of the American Heart Association Stroke Council, Council on Cardiovascular Nursing, Council on Epidemiology and Prevention, Council for High Blood Pressure Research, Council on Peripheral Vascular Disease, and Interdisciplinary Council on Quality of Care and Outcomes Research. Guidelines for the primary prevention of stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2011;42:517–584

7] *AHA Statistical Update Heart Disease and Stroke Statistics—2012 Update. A Report From the American Heart Association Circulation. 2012; 125: e2-e220 Published online before print December 15, 2011, doi: 10.1161/CIR.0b013e31823ac046*

8] David C. Costs and cost-effectiveness of carotid stenting vs. endarterectomy for patients at increased surgical risk: Results from the SAPPHIRE trial. Catheter Cardiovasc Interv. 2011; Mar 1;77(4):463-72

9] Mantese VA, Timaran CH, Chiu D, et al. The Carotid Revascularization Endarterectomy versus Stenting Trial (CREST): stenting versus carotid endarterectomy for carotid disease. Stroke. 2010;41:S31-S34.

10] Mas JL, Trinquart L, Leys D, et al. Endarterectomy Versus Angioplasty in Patients with Symptomatic Severe Carotid Stenosis (EVA-3S) trial: results up to 4 years from a randomised, multicentre trial. Lancet Neurol. 2008;7:885-92.

11] Mast H, Chambless LE, Mohr JP, et al. [Indications for endarterectomy in asymptomatic stenoses of the internal or common carotid artery--results of the North American ACAS Study]. Zentralbl Chir. 1996;121:1033-5.

12] Ringleb PA, Hacke W. [Stent and surgery for symptomatic carotid stenosis. SPACE study results]. Nervenarzt. 2007;78:1130-7.

13] Murad MH, Shahrour A, Shah N, Montori VM. A systematic review and meta-analysis of randomized trials of carotid endarterectomy vs stenting. J Vasc Surg 2011;53:792-7

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
Evidence_Supplement_CARE_Registry_FollowUp_Stroke_Mortality_Assessment_20131218v2.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
This measure is important to determine the number of patients that are being follow-ed after a carotid artery stent procedure. Specifically vital status and if a NIHSS were completed at follow-up.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

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.**

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Many patients after a carotid artery stenting procedure are lost to follow-up.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Frequently performed procedure, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
This is not a PRO-PM.

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 subcriteria 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 : Ischemic Heart Disease, Coronary Artery Disease, Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention : Screening

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.)*

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Measure not previously endorsed

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Neurologic status with an assessment using the NIH Stroke Scale (by an examiner who is certified by the American Stroke Association) AND Discharge Status (alive or Deceased)

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
1 year for numerator and denomiator

S.6. **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, 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.
Field Name: Patient Follow-up Performed Seq No: 9000
Definition: Indicate whether patient follow-up was performed for the procedure. The recommended timeframe for follow-up is 30 days. Occurring between day 21 and the end of day 60 after the procedure. (Days 21-60 inclusive)

1=Yes

Field Name: Follow-Up Date Seq No: 9002
Definition: Indicate the date of follow-up. The recommended timeframe for follow-up is 30 days. Occurring between day 21 and the end of day 60 after the procedure. (Days 21-60 inclusive)

Field Name: Follow Up NIH Stroke Scale Administered Seq No: 9010
Definition: Indicate if the National Institutes of Health Stroke Scale (NIHSS) was administered during follow-up.

1=Yes

Follow-up NIH Stroke Scale Examiner Certified Seq No: 9014
Definition: Indicate the date the National Institutes of Health Stroke Scale (NIHSS) was administered during the follow-up period.
Note - Recommended timeframe to administer NIHSS is within 30 days after the current procedure.
Definition: Indicate if the NIH Stroke Scale examiner who administered the follow-up stroke scale is certified to administer the stroke scale exam. The Stroke Scale assessment should be conducted by someone other than the operator for the current procedure.

1=Yes

Field Name: Follow-up NIH Stroke Scale Examiner Certified Seq No: 9014
Definition: Indicate the date the National Institutes of Health Stroke Scale (NIHSS) was administered during the follow-up period.
Note - Recommended timeframe to administer NIHSS is within 30 days Occurring between day 21 and the end of day 60 after the procedure. (Days 21-60 inclusive)after the current procedure.

Examiner certified= yes
Supporting definitions:
The Stroke Scale assessment should be conducted by someone other than the operator for the current procedure.
Note - NIHSS examiners may become certified through the American Stroke Association.
NIH Stroke Scale Certification is currently available online free of charge: http://learn.heart.org/ihtml/application/student /interface.heart2/nihss.html

Field Name: Patient Status Seq No: 9100
Definition: Indicate if the patient is alive or deceased.

Alive (1) or deceased (2)

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Count of CARE Registry patients that had a carotid artery stenting procedure

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Patients undergoing a carotid artery stent procedure

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Patients deceased at discharge
Patients with was an acute, evolving stroke and dissection

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Field Name: Discharge Status Seq No: 8010
Definition: Indicate whether the patient was alive or deceased at discharge from the hospitalization during which the procedure occurred.
Alive=2

Field Name: Spontaneous Carotid Artery Dissection Seq No: 5060
Definition: Indicate if the patient has had a spontaneous carotid artery dissection prior to the current procedure.

1=Yes

Field Name: Acute Evolving Stroke Seq No: 4340
Definition: Indicate if the patient has experienced an acute evolving stroke with ischemia which is ongoing and progressing at the time of the procedure. Acute evolving stroke includes all of the following:
1. Any sudden development of neurological deficits attributable to cerebral ischemia and/or infarction.
2. Onset of symptoms occurring within prior three days and ongoing at time of procedure.
3. The event is marked by progressively worsening symptoms.
Note: Possible symptoms include, but are not limited to the following: numbness or weakness of the face or body; difficulty speaking or understanding; blurred or decreased vision; dizziness; or loss of balance and coordination.

1=Yes

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
The measure is not stratified.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
No risk adjustment.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)

S.16. Type of score:
Rate/proportion

If other:

S.17. 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.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*

Patient Follow-up is performed and the Follow-up NIH Stroke Scale is administered

The Procedure date of the CAS procedure and the Follow-up Date are equal to or greater than 21 days and less than or equal to 60 days and the follow-up status is deceased or the NIH Stroke Scale is captured.

All patient undergoing a carotid artery stenting, excluding those with an in hospital spontaneous dissection, an acute evolving stroke or discharge status of deceased.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not a PRO=PM measure.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
NCDR Care Registry

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **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 a composite performance measure

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
ACC_NCDR_CARE_Followup_Assessment_Testing_Supplement_101713v2.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: Carotid artery stenting: Evaluation of Vital Status and NIH Stroke Scale at Follow Up
Date of Submission: Click here to enter a date
Type of Measure: Process

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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.

2b2. 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.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present

at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

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.

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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/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*).

We propose to use a clinical registry, the National Cardiovascular Data Registry CARE Registry. This is a national quality improvement registry that is currently participated in by >180 US hospitals. Rigorous quality standards are applied to the data and both quarterly and ad hoc performance reports are generated for participating centers to track and improve their performance.

1.3. What are the dates of the data used in testing? Calendar Year2007Q1-2013Q1

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.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

A cohort of the NCDR CARE Registry (2007Q1 and 2013 Q1) was used to establish the prevalence of neurological function testing. We restricted our analyses to those hospitals that performed 30 or more carotid revascularization procedures to improve the precision of our estimates. To examine the test-restest validity of the measured data elements, we expanded the time window of CARE from 2007-2013 to identify patients with 2 or more procedures.

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

We included a total of 18,212 patients. The process measure we assessed was vital status and the presences of an NIH Stroke Scale (NIHSS) assessment at 30 days after their procedure. The only exclusion for this measure would be patients who are acutely having a stroke at the time of treatment, as it is not feasible to distinguish whether or not 30-day neurological outcomes were due to the presenting stroke, or treatment. The characteristics of the patients, stratified by collection of their outcomes data, is provided below:

	Total	Both Vital and NIHSS at F/U		
	n = 18212	1 n = 7220	0 n = 10992	P-Value
Followup Measures				
Has_vital	12883 (70.74%)	7220 (100.00%)	5663 (51.52%)	< 0.001
has_nihss	7220 (39.64%)	7220 (100.00%)	0 (0.00%)	< 0.001
A. Demographics				
age	70.55 ± 10.43	70.87 ± 10.06	70.33 ± 10.65	< 0.001
Sex Male Female Missing (.)	 11265 (61.86%) 6946 (38.14%) 1	 4540 (62.88%) 2680 (37.12%)	 6725 (61.19%) 4266 (38.81%) 1	0.021
Race White Black/African American Asian American Indian/Alaskan Native Native Hawaiian/Pacific Islander Other Missing (.)	 16716 (91.88%) 853 (4.69%) 128 (0.70%) 58 (0.32%) 15 (0.08%) 423 (2.33%) 19	 6702 (92.89%) 342 (4.74%) 45 (0.62%) 8 (0.11%) 3 (0.04%) 115 (1.59%) 5	 10014 (91.22%) 511 (4.65%) 83 (0.76%) 50 (0.46%) 12 (0.11%) 308 (2.81%) 14	< 0.001
Preprocedure Creatinine Level Missing	1.18 ± 0.73 593	1.19 ± 0.72 148	1.18 ± 0.74 445	0.272
Currently On Dialysis Missing (.)	468 (2.58%) 68	154 (2.14%) 9	314 (2.87%) 59	0.002
Tobacco History Current Former Never Missing (.)	 5066 (28.01%) 8350 (46.17%) 4669 (25.82%) 127	 1993 (27.69%) 3495 (48.56%) 1710 (23.76%) 22	 3073 (28.23%) 4855 (44.59%) 2959 (27.18%) 105	< 0.001

	Total	Both Vital and NIHSS at F/U		
	n = 18212	1 n = 7220	0 n = 10992	P-Value
Hypertension Missing (.)	16437 (90.53%) 55	6612 (91.59%) 1	9825 (89.82%) 54	< 0.001
Dyslipidemia Missing (.)	15793 (86.99%) 58	6594 (91.34%) 1	9199 (84.12%) 57	< 0.001
Peripheral Arterial Disease Missing (.)	7628 (42.04%) 69	3060 (42.39%) 2	4568 (41.81%) 67	0.437
Diabetes Mellitus Missing (.)	6823 (37.59%) 63	2774 (38.43%) 2	4049 (37.04%) 61	0.058
Ischemic Heart Disease Missing (.)	9732 (53.63%) 65	4022 (55.74%) 4	5710 (52.24%) 61	< 0.001
History of Heart Failure Missing (.)	2997 (16.52%) 71	1229 (17.03%) 3	1768 (16.18%) 68	0.134
Most Recent LVEF% Missing	53.25 ± 13.75 8279	53.08 ± 13.82 3081	53.38 ± 13.71 5198	0.286
History of Atrial Fibrillation or Flutter Missing (.)	2299 (12.68%) 87	916 (12.70%) 10	1383 (12.67%) 77	0.946
Restenosis in Target Vessel After Prior CAS Missing (.)	568 (3.13%) 69	166 (2.30%) 6	402 (3.68%) 63	< 0.001
Restenosis in Target Vessel After Prior CEA Missing (.)	2800 (15.43%) 70	1023 (14.18%) 5	1777 (16.26%) 65	< 0.001
Target Lesion Symptomatic w/in Past 6 Months Missing (.)	7690 (42.40%) 77	2254 (31.28%) 13	5436 (49.74%) 64	< 0.001
Aortic Arch Type Type I Type II Type III Missing (.)	 8511 (51.60%) 6290 (38.14%) 1692 (10.26%) 1719	 3383 (49.15%) 2710 (39.37%) 790 (11.48%) 337	 5128 (53.36%) 3580 (37.25%) 902 (9.39%) 1382	< 0.001
Bovine Arch Missing (.)	2330 (13.63%) 1119	992 (14.21%) 238	1338 (13.23%) 881	0.068
Lesion Location High Cervical Low Intrathoracic Missing (.) N	 2105 (83.96%) 402 (16.04%) 578 15127	 828 (87.25%) 121 (12.75%) 91 6180	 1277 (81.96%) 281 (18.04%) 487 8947	< 0.001
Visible Thrombus Present Missing (.)	18212	7220	10992	
CAS only elements				
Visible Thrombus Present Missing (.)	761 (4.25%) 297	222 (3.10%) 66	539 (5.01%) 231	< 0.001
Ulceration Missing (.)	4937 (27.63%) 341	2254 (31.51%) 67	2683 (25.03%) 274	< 0.001
Calcification None Mild to Moderate Dense and Concentric Missing (.)	 6792 (38.18%) 8575 (48.21%) 2421 (13.61%) 424	 2207 (30.85%) 4091 (57.18%) 857 (11.98%) 65	 4585 (43.12%) 4484 (42.17%) 1564 (14.71%) 359	< 0.001
Lesion Length Missing	19.91 ± 10.63 2030	19.10 ± 9.02 264	20.53 ± 11.66 1766	< 0.001
Minimum Luminal Diameter (MLD) Missing	1.79 ± 2.04 3445	1.55 ± 1.82 972	1.96 ± 2.17 2473	< 0.001
Diameter of Distal (non-tapered) ICA for NASCET Missing	5.70 ± 1.70 3542	5.59 ± 1.44 897	5.78 ± 1.87 2645	< 0.001

	Total	Both Vital and NIHSS at F/U		
	n = 18212	1 n = 7220	0 n = 10992	P-Value
Preprocedure % Stenosis Missing	84.27 ± 11.36 263	84.45 ± 10.31 24	84.14 ± 12.02 239	0.073
Lesion Treatment Incomplete or Aborted Missing (.)	168 (0.93%) 140	27 (0.38%) 22	141 (1.30%) 118	
Postdilation Performed Missing (.)	16102 (88.88%) 95	6558 (90.98%) 12	9544 (87.49%) 83	< 0.001
Nominal Balloon Diameter Missing	5.43 ± 1.40 2345	5.39 ± 1.37 702	5.45 ± 1.43 1643	0.005
Maximum Inflation Pressure Missing	9.84 ± 3.18 2739	10.13 ± 3.44 763	9.63 ± 2.97 1976	< 0.001
Final Minimum Luminal Diameter Missing	5.38 ± 1.64 3748	5.27 ± 1.51 1130	5.47 ± 1.72 2618	< 0.001
Final % Stenosis Missing	8.60 ± 11.94 359	8.56 ± 10.43 27	8.62 ± 12.86 332	0.729
Neuroligic History and Risk Factors Pre procedure				
Dementia or Alzheimer s Disease Missing (.)	592 (3.26%) 52	184 (2.55%) 1	408 (3.73%) 51	< 0.001
History of Seizure or Known Seizure Disorder Missing (.)	484 (2.67%) 53	166 (2.30%) 3	318 (2.91%) 50	0.013
Neurologic Event(s) Prior to Procedure Missing (.)	8852 (48.77%) 62	2958 (41.00%) 5	5894 (53.90%) 57	< 0.001
Prior TIA	5707 (31.34%)	1919 (26.58%)	3788 (34.46%)	< 0.001
Prior Ischemic stroke	2988 (16.41%)	985 (13.64%)	2003 (18.22%)	< 0.001
Prior Hemorrhage or Hemorrhagic Stroke	112 (0.61%)	27 (0.37%)	85 (0.77%)	< 0.001
Acute Evolving Stroke Missing (.)	523 (2.89%) 112	117 (1.62%) 15	406 (3.73%) 97	< 0.001
Neurologic Status Preprocedure				
Preprocedure NIH Stroke Scale Total Score Missing	0.94 ± 2.77 5330	0.61 ± 1.79 277	1.33 ± 3.55 5053	< 0.001
Preprocedure Modified Rankin Score Missing	0.51 ± 0.97 10420	0.35 ± 0.80 2760	0.72 ± 1.13 7660	< 0.001
CEA and CAS elements				
Fluoro Time Missing	17.92 ± 14.63 674	16.29 ± 12.06 110	19.03 ± 16.06 564	< 0.001
Procedural Arterial Access Site Femoral Brachial/Radial/Axillary Direct Carotid Puncture Carotid Cutdown Other Missing (.)	 17702 (97.97%) 279 (1.54%) 37 (0.20%) 45 (0.25%) 6 (0.03%) 143	 6975 (97.01%) 201 (2.80%) 11 (0.15%) 3 (0.04%) 0 (0.00%) 30	 10727 (98.60%) 78 (0.72%) 26 (0.24%) 42 (0.39%) 6 (0.06%) 113	< 0.001
Pre Procedural Meds				
ASA before procedure No Yes Contraindicated Missing (.)	 1875 (10.33%) 16146 (88.94%) 133 (0.73%) 58	 380 (5.27%) 6790 (94.11%) 45 (0.62%) 5	 1495 (13.67%) 9356 (85.53%) 88 (0.80%) 53	< 0.001

	Total	Both Vital and NIHSS at F/U		
	n = 18212	1 n = 7220	0 n = 10992	P-Value
CLOPIDOGREL before procedure				< 0.001
No	2720 (14.98%)	710 (9.84%)	2010 (18.37%)	
Yes	15328 (84.43%)	6472 (89.70%)	8856 (80.96%)	
Contraindicated	106 (0.58%)	33 (0.46%)	73 (0.67%)	
Missing (.)	58	5	53	
TICLOPIDINE before procedure				0.047
No	17880 (98.62%)	7098 (98.43%)	10782 (98.75%)	
Yes	173 (0.95%)	72 (1.00%)	101 (0.92%)	
Contraindicated	77 (0.42%)	41 (0.57%)	36 (0.33%)	
Missing (.)	82	9	73	
Intra Procedure Meds				
UNFRACTIONATED HEPARIN during procedure				< 0.001
No	7178 (39.57%)	3599 (49.91%)	3579 (32.74%)	
Yes	10941 (60.31%)	3604 (49.98%)	7337 (67.12%)	
Contraindicated	23 (0.13%)	8 (0.11%)	15 (0.14%)	
Missing (.)	70	9	61	
LMWH during procedure				< 0.001
No	17705 (97.72%)	7077 (98.17%)	10628 (97.42%)	
Yes	374 (2.06%)	103 (1.43%)	271 (2.48%)	
Contraindicated	39 (0.22%)	29 (0.40%)	10 (0.09%)	
Missing (.)	94	11	83	
ANY_THROMBININHIBITORS during procedure				< 0.001
No	11562 (63.81%)	3760 (52.16%)	7802 (71.51%)	
Yes	6527 (36.02%)	3426 (47.53%)	3101 (28.42%)	
Contraindicated	30 (0.17%)	22 (0.31%)	8 (0.07%)	
Missing (.)	93	12	81	
Post Procedure Meds				
UNFRACTIONATED HEPARIN				< 0.001
No	17250 (95.26%)	7007 (97.21%)	10243 (93.97%)	
Yes	818 (4.52%)	176 (2.44%)	642 (5.89%)	
Contraindicated	40 (0.22%)	25 (0.35%)	15 (0.14%)	
Missing (.)	104	12	92	
LMWH				< 0.001
No	17577 (97.09%)	7083 (98.27%)	10494 (96.31%)	
Yes	491 (2.71%)	100 (1.39%)	391 (3.59%)	
Contraindicated	36 (0.20%)	25 (0.35%)	11 (0.10%)	
Missing (.)	108	12	96	
Discharge Meds				
ASA at discharge				< 0.001
No	1140 (6.29%)	290 (4.02%)	850 (7.78%)	
Yes	16840 (92.84%)	6861 (95.13%)	9979 (91.33%)	
Contraindicated	158 (0.87%)	61 (0.85%)	97 (0.89%)	
Missing (.)	74	8	66	
CLOPIDOGREL at discharge				< 0.001
No	953 (5.25%)	300 (4.16%)	653 (5.97%)	
Yes	17078 (94.14%)	6870 (95.26%)	10208 (93.40%)	
Contraindicated	110 (0.61%)	42 (0.58%)	68 (0.62%)	
Missing (.)	71	8	63	

	Total	Both Vital and NIHSS at F/U		
	n = 18212	1 n = 7220	0 n = 10992	P-Value
TICLOPIDINE at discharge				0.051
No	17848 (98.60%)	7083 (98.35%)	10765 (98.76%)	
Yes	185 (1.02%)	84 (1.17%)	101 (0.93%)	
Contraindicated	69 (0.38%)	35 (0.49%)	34 (0.31%)	
Missing (.)	110	18	92	
WARFARIN at discharge				0.132
No	16586 (91.61%)	6630 (92.03%)	9956 (91.33%)	
Yes	1444 (7.98%)	541 (7.51%)	903 (8.28%)	
Contraindicated	75 (0.41%)	33 (0.46%)	42 (0.39%)	
Missing (.)	107	16	91	

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.

We used 2012 data to assess the prevalence of complete follow-up. We used all available data from 3007-2013 to assess the test-retest reliability of the data elements used to describe patient characteristics. We also restricted this test-retest cohort to those hospitals that performed >30 procedures over this time period so as to provide more reliable estimates of the reproducibility of these data elements.

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*)

Level	Signal-to-Noise
All, >10 Procedures	.984
>Q1 (>55 Procedures)	.986
>Q2 (>91 Procedures)	.989
>Q3 (>161 Procedures)	.993
>Average (>131 Procedures)	.992

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*)

2a2.3. For each level 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)

We compared, among centers reporting greater than 30 cases (n=XX) between 2007-2013, the accuracy of data elements entered into the CARE registry. This approach enabled us to examine 2 independent abstractions of data for the same patient. For certain characteristics that would not change (e.g. gender), we would expect near perfect reproducibility. For other characteristics (e.g. diabetes) we would expect that any patient diagnosed with diabetes on the first visit should also have diabetes recorded on the second visit. It is, however, clinically plausible that someone could be diagnosed with diabetes between their first and second visit, so the emergence of diabetes on the second visit is not necessarily an 'error' and no interpretation is made for these scenarios.

There were 449 patients in the CARE registry that had 2 procedures between 2007-2013. Important data elements, support the overall validity of the registry, are provided below:

Age, as defined by date of birth, did not differ in any of the cases.

Gender did not vary in any of the patient records.

Race did not vary in any of the records.

Smoking had minimal inconsistencies. There were 3 patients (0.67%) who were categorized as current smokers on their 1st procedure and never smokers on their 2nd procedure. There were 9 patients (2.0%) listed as former smokers on their 1st visit and never smokers on their 2nd procedure.

History of Peripheral Artery Disease was noted in 11 patients (2.4%) at the time of their 1st procedure, but not at the time of their 2nd.

Diabetes was noted in 6 patients (1.3%) at the time of their 1st procedure, but not at the time of their 2nd.

Chronic Lung Disease was noted in 7 patients (1.6%) at the time of their 1st procedure, but not at the time of their 2nd.

Dyslipidemia was noted in 12 patients (2.7%) at the time of their 1st procedure, but not at the time of their 2nd.

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?*)

For the elements that we were able to assess, we believe that the results reported in the NCDR CARE record are consistent and accurate. There was not an independent audit performed for the CARE registry, but we observed <3% 'mistakes' in 1 of the records for patients who had 2 procedures reported in the CARE Registry.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

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

☐ **Performance measure score**

- ☐ **Empirical validity testing**
- ☐ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or

resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level 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)*

Content validity of this outcome – post-procedure stroke is one of the major adverse outcomes from carotid artery stenting and carotid endarterectomy. For example, this was the major outcome in the recent CREST trial, a randomized comparison of carotid artery stenting and surgical endarterectomy. A recent meta-analysis by Bangalore and colleagues (Arch Nerol 201; 68:172-84) summarized the results of 13 randomized controlled trials to assess the comparative effectiveness of stenting vs. surgery for carotid occlusion on the outcomes of myocardial infarction, death and stroke. The latter 2 of these outcomes are proposed to serve as a process measure for quantifying the quality of carotid revascularization by this measure.

There is a very sound clinical rationale for systematically measuring the outcomes of carotid revascularization. First, without knowing the outcomes, a hospital cannot know if it is applying its treatment in a safe and effective manner. Given how infrequently current providers assess the 30-day survival and stroke outcomes, it is obvious that more than half of these hospitals have no foundation with which to assess the quality of their care. We have proposed a process measure, merely assessing the stroke-free survival of treated patients, because without clearer ascertainment of outcomes it is not possible to provide risk-adjusted comparisons across centers and provide clear benchmarks of performance to identify hospitals that have the opportunity to improve. Second, as the country seeks to support the use of evidence-based medicine, the majority of the evidence in carotid disease comes from clinical trials. However, many of the trials establishing the benefits of carotid revascularization require that centers document a certain success rate, without complications of stroke or death, before the center can participate in a clinical trial. If a center does not know its rate, it will not know whether or not the benefits observed in a clinical trial apply to their practice. Finally, for patients to be adequately informed about the risks and benefits of treatment, hospitals need to have reliable data to share with their patients. By collecting, analyzing and reporting the outcomes of treatment, hospitals will be much better able to provide their patients the information that they need to make a treatment decision.

Face Validity of this outcome- As expressed in the application, this measure was shared with a noted panel of expert cardiologist who participate regularly in carotid artery stenting procedures. All members of the committee reported that the measures appears to be a good indication of a positive outcome in carotid artery stenting procedures.

In developing this measure, the ACC consulted with leading experts in the field and vetted the process measure with the following committees. The individuals within specific committees and workgroups are noted below:

NCDR Strategic Quality and Oversight Committee— an ACC leadership oversight committee that serves as the primary resource for crosscutting scientific and quality of care methodological issues – ensured the data dictionaries and metrics are consistent across registries. They also reviewed and approved the methodology and results of the bleeding outcome and model.
These members include Dr. Frederick Masoudi (chair) , Dr. David Malenka, Dr. Thomas Tsai, Dr. Matthew Reynolds,Dr. David Shahian, Dr. John Windle, Dr. Fred Resnic, Dr. John Moore, Dr. Deepak Bhatt, Dr. James Tcheng, Dr. Jeptha Curtis, Dr. Paul Chan, Dr. Matt Roe, and Dr. John Rumsfeld

NCDR Clinical SubWorkgroup is a designated set of experts that oversees this NQF application. Prior to submission, it ensures there is variation in care, disparities data, and that the measure is a true reflection of quality care at a particular site and can also be used to improve quality. This committee included Dr. Jeptha Curtis (chair), Dr. Frederick Masoudi, Dr. John Rumsfeld, Dr. Christopher White, and Dr. Thomas Tsai.

NCDR CARE/PVI Transition Committee provides strategic direction for the Registry and ensures the measures submitted to NQF met key criterion such as reliability, feasibility, and that there is compelling evidence base behind the development and implementation of this measure, which included Christopher White (Chair), Kalon Ho, Ken Rosenfield, Bobby Yeh, Michael Jaff, Thomas Tsai, P. Michael Grossman, Herbert Aronow, H. Vernon Anderson

Lastly the 16 member NCDR Management Board and 31member ACCF Board of Trustees approved these measures for submission to NQF.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

As this measure is being proposed primarily on the basis of its content validity, as described above, there are no empiric results from formal validity testing.

2b2.4. 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?*)

As described above, we believe that acquiring the short-term outcomes of carotid revascularization is a critical foundation for assessing and improving care. The CARE registry provides the infrastructure to enter and analyze these data, if collected. An approved performance measure will increase the acquisition of these data and enable quality to be assessed and improved.

_____

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

The only proposed exclusion is for patients being treated in the context of an acute evolving stroke. This is a distinct clinical setting from the treatment of stable carotid disease. Moreover, the neurological results are likely to be strongly influenced by the presenting stroke, more so than the revascularization procedure that they receive.

2b3.2. 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*)

The Table below shows the differences in clinical characteristics of those with an acute evolving stroke as compared with those treated for stable carotid disease:

ex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	16616	91.24	16616	91.24
AES	491	2.70	17107	93.93
Spont Dis	182	1.00	17289	94.93
Hosp <30 procs	923	5.07	18212	100.00

2b3.3. 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*)

Given the 2 very distinct populations, we believe that a measure of the survival and neurological outcomes at 30 days is an internally consistent, clinically-interpretable measure that does not suffer from excluding those with an acute ischemic stroke.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

Not applicable.

2b4.4. What were the statistical results of the analyses used to select risk factors?

Not applicable.

2b4.5. 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*)

Not applicable.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

Not applicable.

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable.

2b4.9. Results of Risk Stratification Analysis:

Not applicable.

2b4.10. 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*)

Not applicable.

***2b4.11. 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*)

Not applicable.

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. 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)*

We observed marked variation in the collection of 30-day outcomes data across hospitals. Among 180 hospitals performing carotid revascularization in 17,289 patients in 2012, the range of hospital's collection of 30-day outcomes varied from 0% to 100%. Hospitals with <30 procedures were excluded. The interquartile ranges were 0-3.2%, 3.2-26.6%, 26.6-59.3% and 59.3-100%. The variation in patient characteristics, by quartile, is provided below:

	Total	Both Vital and NIHSS at F/U		
	n = 16616	1 n = 6833	0 n = 9783	P-Value
Followup Measures				
Has_vital	11899 (71.61%)	6833 (100.00%)	5066 (51.78%)	< 0.001
has_nihss	6833 (41.12%)	6833 (100.00%)	0 (0.00%)	< 0.001
A. Demographics				
age	70.69 ± 10.24	70.84 ± 10.02	70.59 ± 10.39	0.122
Sex Male Female Missing (.)	 10270 (61.81%) 6345 (38.19%) 1	 4291 (62.80%) 2542 (37.20%)	 5979 (61.12%) 3803 (38.88%) 1	0.029
Race White Black/African American Asian American Indian/Alaskan Native Native Hawaiian/Pacific Islander Other Missing (.)	 15268 (91.99%) 765 (4.61%) 113 (0.68%) 54 (0.33%) 12 (0.07%) 386 (2.33%) 18	 6344 (92.91%) 321 (4.70%) 40 (0.59%) 8 (0.12%) 3 (0.04%) 112 (1.64%) 5	 8924 (91.34%) 444 (4.54%) 73 (0.75%) 46 (0.47%) 9 (0.09%) 274 (2.80%) 13	< 0.001
Preprocedure Creatinine Level Missing	1.19 ± 0.74 536	1.19 ± 0.71 138	1.19 ± 0.76 398	0.721
Currently On Dialysis Missing (.)	425 (2.57%) 65	145 (2.12%) 9	280 (2.88%) 56	0.003
Tobacco History Current Former Never Missing (.)	 4611 (27.94%) 7679 (46.53%) 4213 (25.53%) 113	 1890 (27.75%) 3319 (48.72%) 1603 (23.53%) 21	 2721 (28.08%) 4360 (44.99%) 2610 (26.93%) 92	< 0.001
Hypertension Missing (.)	15112 (91.24%) 53	6279 (91.91%) 1	8833 (90.77%) 52	0.011
Dyslipidemia Missing (.)	14566 (87.95%) 55	6265 (91.70%) 1	8301 (85.32%) 54	< 0.001
Peripheral Arterial Disease Missing (.)	7030 (42.47%) 64	2912 (42.63%) 2	4118 (42.36%) 62	0.732
Diabetes Mellitus Missing (.)	6271 (37.88%) 59	2637 (38.60%) 2	3634 (37.36%) 57	0.105
Ischemic Heart Disease Missing (.)	8991 (54.31%) 60	3818 (55.90%) 3	5173 (53.19%) 57	< 0.001
History of Heart Failure Missing (.)	2768 (16.73%) 66	1168 (17.10%) 3	1600 (16.46%) 63	0.277
Most Recent LVEF% Missing	53.18 ± 13.76 7533	53.08 ± 13.76 2919	53.24 ± 13.77 4614	0.579
History of Atrial Fibrillation or Flutter Missing (.)	2100 (12.70%) 82	868 (12.72%) 9	1232 (12.69%) 73	0.952
Restenosis in Target Vessel After Prior CAS Missing (.)	519 (3.14%) 65	155 (2.27%) 6	364 (3.74%) 59	< 0.001

	Total	Both Vital and NIHSS at F/U		
	n = 16616	1 n = 6833	0 n = 9783	P-Value
Restenosis in Target Vessel After Prior CEA Missing (.)	2517 (15.21%) 65	948 (13.88%) 5	1569 (16.14%) 60	< 0.001
Target Lesion Symptomatic w/in Past 6 Months Missing (.)	6799 (41.10%) 73	2087 (30.60%) 12	4712 (48.47%) 61	< 0.001
Aortic Arch Type Type I Type II Type III Missing (.)	 7728 (51.17%) 5807 (38.45%) 1569 (10.39%) 1512	 3176 (48.76%) 2576 (39.55%) 761 (11.68%) 320	 4552 (52.99%) 3231 (37.61%) 808 (9.41%) 1192	< 0.001
Bovine Arch Missing (.)	2149 (13.76%) 1000	949 (14.37%) 227	1200 (13.32%) 773	0.061
Lesion Location High Cervical Low Intrathoracic Missing (.) N	 1841 (83.91%) 353 (16.09%) 522 13900	 755 (86.88%) 114 (13.12%) 84 5880	 1086 (81.96%) 239 (18.04%) 438 8020	0.002
Visible Thrombus Present Missing (.)	 16616	 6833	 9783	
CAS only elements				
Visible Thrombus Present Missing (.)	562 (3.44%) 273	196 (2.90%) 64	366 (3.82%) 209	0.001
Ulceration Missing (.)	4571 (28.02%) 301	2140 (31.61%) 63	2431 (25.47%) 238	< 0.001
Calcification 0 None Mild to Moderate Missing (.)	 6117 (37.65%) 7912 (48.70%) 2218 (13.65%) 369	 2081 (30.72%) 3874 (57.20%) 818 (12.08%) 60	 4036 (42.60%) 4038 (42.62%) 1400 (14.78%) 309	< 0.001
Lesion Length Missing	19.77 ± 10.10 1772	19.02 ± 8.89 252	20.36 ± 10.93 1520	< 0.001
Minimum Luminal Diameter (MLD) Missing	1.75 ± 1.99 3122	1.53 ± 1.80 938	1.92 ± 2.11 2184	< 0.001
Diameter of Distal (non-tapered) ICA for NASCET Missing	5.71 ± 1.64 3187	5.59 ± 1.42 870	5.80 ± 1.80 2317	< 0.001
Preprocedure % Stenosis Missing	84.26 ± 10.92 223	84.39 ± 10.22 21	84.16 ± 11.39 202	0.192
Lesion Treatment Incomplete or Aborted Missing (.)	148 (0.90%) 128	24 (0.35%) 20	124 (1.28%) 108	
Reasons Treatment Aborted Missing (.)	142 (100.00%) 16474	24 (100.00%) 6809	118 (100.00%) 9665	
Postdilation Performed Missing (.)	14839 (89.76%) 85	6228 (91.29%) 11	8611 (88.69%) 74	< 0.001
Nominal Balloon Diameter Missing	5.42 ± 1.38 1964	5.38 ± 1.33 643	5.45 ± 1.41 1321	0.002
Maximum Inflation Pressure Missing	9.84 ± 3.16 2329	10.12 ± 3.42 700	9.63 ± 2.93 1629	< 0.001
Final Minimum Luminal Diameter Missing	5.39 ± 1.61 3399	5.26 ± 1.46 1084	5.48 ± 1.70 2315	< 0.001
Final % Stenosis Missing	8.57 ± 11.63 294	8.55 ± 10.22 24	8.59 ± 12.54 270	0.852
Neuroligic History and Risk Factors Preprocedure				

	Total	Both Vital and NIHSS at F/U		
	n = 16616	1 n = 6833	0 n = 9783	P-Value
Dementia or Alzheimer s Disease Missing (.)	521 (3.14%) 50	164 (2.40%) 1	357 (3.67%) 49	< 0.001
History of Seizure or Known Seizure Disorder Missing (.)	432 (2.61%) 52	155 (2.27%) 3	277 (2.85%) 49	0.022
Neurologic Event(s) Prior to Procedure Missing (.)	7865 (47.50%) 59	2755 (40.34%) 4	5110 (52.53%) 55	< 0.001
Prior TIA	5187 (31.22%)	1798 (26.31%)	3389 (34.64%)	< 0.001
Prior Ischemic stroke	2571 (15.47%)	908 (13.29%)	1663 (17.00%)	< 0.001
Prior Hemorrhage or Hemorrhagic Stroke	92 (0.55%)	23 (0.34%)	69 (0.71%)	0.002
Acute Evolving Stroke Missing (.)	0 (0.00%) 100	0 (0.00%) 14	0 (0.00%) 86	
Neurologic Status Preprocedure				
Preprocedure NIH Stroke Scale Total Score Missing	0.72 ± 2.06 4718	0.55 ± 1.54 245	0.93 ± 2.55 4473	< 0.001
Preprocedure Modified Rankin Score Missing	0.48 ± 0.92 9325	0.34 ± 0.78 2591	0.67 ± 1.07 6734	< 0.001
CEA and CAS elements				
Fluoro Time Missing	17.34 ± 13.67 596	15.98 ± 11.70 106	18.33 ± 14.86 490	< 0.001
Procedural Arterial Access Site Femoral Brachial/Radial/Axillary Direct Carotid Puncture Carotid Cutdown Other Missing (.)	 16135 (97.88%) 273 (1.66%) 31 (0.19%) 40 (0.24%) 5 (0.03%) 132	 6590 (96.84%) 201 (2.95%) 11 (0.16%) 3 (0.04%) 0 (0.00%) 28	 9545 (98.62%) 72 (0.74%) 20 (0.21%) 37 (0.38%) 5 (0.05%) 104	< 0.001
Pre Procedural Meds				
ASA_Pre No Yes Contra Missing (.)	 1571 (9.49%) 14877 (89.83%) 113 (0.68%) 55	 328 (4.80%) 6458 (94.58%) 42 (0.62%) 5	 1243 (12.77%) 8419 (86.50%) 71 (0.73%) 50	< 0.001
CLOPIDOGREL_Pre No Yes Contra Missing (.)	 2357 (14.23%) 14113 (85.22%) 91 (0.55%) 55	 643 (9.42%) 6153 (90.11%) 32 (0.47%) 5	 1714 (17.61%) 7960 (81.78%) 59 (0.61%) 50	< 0.001
TICLOPIDINE_Pre No Yes Contra Missing (.)	 16313 (98.59%) 166 (1.00%) 68 (0.41%) 69	 6715 (98.39%) 70 (1.03%) 40 (0.59%) 8	 9598 (98.72%) 96 (0.99%) 28 (0.29%) 61	0.012
Intra Procedure Meds				
UNFRACTIONATED HEPARIN_intra No Yes Contra Missing (.)	 6795 (41.06%) 9735 (58.83%) 19 (0.11%) 67	 3467 (50.81%) 3350 (49.09%) 7 (0.10%) 9	 3328 (34.22%) 6385 (65.66%) 12 (0.12%) 58	< 0.001
LMWH_intra No Yes Contra Missing (.)	 16143 (97.64%) 352 (2.13%) 38 (0.23%) 83	 6697 (98.17%) 97 (1.42%) 28 (0.41%) 11	 9446 (97.27%) 255 (2.63%) 10 (0.10%) 72	< 0.001

	Total	Both Vital and NIHSS at F/U		
	n = 16616	1 n = 6833	0 n = 9783	P-Value
ANY_THROMBININHIBITORS_intra				< 0.001
No	10293 (62.26%)	3485 (51.09%)	6808 (70.11%)	
Yes	6210 (37.57%)	3314 (48.59%)	2896 (29.82%)	
Contra	28 (0.17%)	22 (0.32%)	6 (0.06%)	
Missing (.)	85	12	73	
Post Procedure Meds				
UNFRACTIONATED HEPARIN_Post				< 0.001
No	15829 (95.80%)	6644 (97.41%)	9185 (94.67%)	
Yes	658 (3.98%)	153 (2.24%)	505 (5.21%)	
Contra	36 (0.22%)	24 (0.35%)	12 (0.12%)	
Missing (.)	93	12	81	
LMWH_Post				< 0.001
No	16103 (97.47%)	6706 (98.31%)	9397 (96.88%)	
Yes	385 (2.33%)	91 (1.33%)	294 (3.03%)	
Contra	33 (0.20%)	24 (0.35%)	9 (0.09%)	
Missing (.)	95	12	83	
Discharge Meds				
ASA_dc				< 0.001
No	966 (5.84%)	260 (3.81%)	706 (7.26%)	
Yes	15447 (93.31%)	6509 (95.34%)	8938 (91.89%)	
Contra	141 (0.85%)	58 (0.85%)	83 (0.85%)	
Missing (.)	62	6	56	
CLOPIDOGREL_dc				< 0.001
No	828 (5.00%)	273 (4.00%)	555 (5.71%)	
Yes	15628 (94.40%)	6514 (95.42%)	9114 (93.69%)	
Contra	99 (0.60%)	40 (0.59%)	59 (0.61%)	
Missing (.)	61	6	55	
TICLOPIDINE_dc				0.017
No	16300 (98.61%)	6705 (98.34%)	9595 (98.81%)	
Yes	171 (1.03%)	80 (1.17%)	91 (0.94%)	
Contra	58 (0.35%)	33 (0.48%)	25 (0.26%)	
Missing (.)	87	15	72	
WARFARIN_dc				0.072
No	15140 (91.59%)	6279 (92.07%)	8861 (91.26%)	
Yes	1324 (8.01%)	510 (7.48%)	814 (8.38%)	
Contra	66 (0.40%)	31 (0.45%)	35 (0.36%)	
Missing (.)	86	13	73	

The variations in hospital characteristics, stratified by quartile of follow-up data collection, is shown below:

	Total	Rate of Collecting 30-day Survival and Neurological Status by Quartile				
	n = 126	Quartile 1 (0 to <3.2%) n = 31	Quartile 2 (3.2 to <26.5%) n = 32	Quartile 3 (26.5 to <59.3%) n = 31	Quartile 4 (59.3 to 100%) n = 32	P-Value
ishTeaching Hospital	56 (44.44%)	12 (38.71%)	17 (53.13%)	14 (45.16%)	13 (40.63%)	0.663
Public Hospital	68 (53.97%)	15 (48.39%)	16 (50.00%)	17 (54.84%)	20 (62.50%)	0.673
Type of Hospital						0.663
Government	3 (2.38%)	1 (3.23%)	2 (6.25%)	0 (0.00%)	0 (0.00%)	
Private/Community	113 (89.68%)	28 (90.32%)	27 (84.38%)	29 (93.55%)	29 (90.63%)	
University	10 (7.94%)	2 (6.45%)	3 (9.38%)	2 (6.45%)	3 (9.38%)	

	Total	Rate of Collecting 30-day Survival and Neurological Status by Quartile				
	n = 126	Quartile 1 (0 to <3.2%) n = 31	Quartile 2 (3.2 to <26.5%) n = 32	Quartile 3 (26.5 to <59.3%) n = 31	Quartile 4 (59.3 to 100%) n = 32	P-Value
Location of Hospital						0.640
Rural	20 (15.87%)	6 (19.35%)	3 (9.38%)	6 (19.35%)	5 (15.63%)	
Suburban	49 (38.89%)	13 (41.94%)	15 (46.88%)	8 (25.81%)	13 (40.63%)	
Urban	57 (45.24%)	12 (38.71%)	14 (43.75%)	17 (54.84%)	14 (43.75%)	
Procedural Volume in 2012	137.21 ± 140.31	92.61 ± 81.64	124.53 ± 88.36	128.00 ± 81.16	202.03 ± 228.40	0.014

2b5.2. 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*)

We observed that patient characteristics of hospitals in the lower performance quartiles or outcomes reporting did not differ substantially from a clinical perspective, other than that their patients were more likely to be symptomatic within the past 6 months with worse NIHSS and modified Rankin Scores pre-procedurally and that they trended to treat larger vessels when performing stenting. We also noted a strong association between procedural volume and follow-up rates, with the average procedural volume at the best performing hospitals being substantially greater than poorer performing sites.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

The acquisition of outcomes data by hospitals after performing carotid revascularization could not be more broad ranging, from never to always assessing patients' outcomes. As described in Section 2b2.2, knowing and hospital's performance is essential for providing safe, evidence-based, patient-centered care. Importantly, since some hospitals understanding a were able to assess the survival and neurological outcomes of all of their patients, it is currently feasible to do so.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
***If only one set of specifications, this section can be skipped*.**

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different datasources/specifications (*describe the steps—do not just name a*

method; what statistical analysis was used)

We are not proposing alternative methods for data collection or performance assessment.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
If other:

3b. Electronic Sources

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

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
No feasibility assessment **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. Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Participating hospitals report patient demographics, medical history, risk factors, hospital presentation, initial cardiac status, procedural details, medications, laboratory values and in-hospital outcomes. The majority of the required data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions can manually report using a free web-based tool or automate the reporting by using certified software developed by third-party vendors. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

The NCDR Data Quality Program consists of 3 main components: data completeness, consistency, and accuracy. Completeness focuses on the proportion of missing data within fields, whereas consistency determines the extent to which logically related fields contain values consistent with other fields. Accuracy characterizes the agreement between registry data and the contents of original charts from the hospitals submitting data.

The Data Quality Report (DQR) consists of registry-specific algorithms that require predetermined levels of completeness and consistency for submitted data fields. Before entering the Enterprise Data Warehouse (EDW), all submissions are scored for file integrity and data completeness, receiving 1 of 3 scores that are transmitted back to facilities using a color coding scheme. A "red light" means that a submission has failed because of file integrity problems such as excessive missing data and internally inconsistent data. Such data are not processed or loaded into the EDW. A "yellow light" status means that a submission has passed the integrity checks but failed in completeness according to predetermined thresholds. Such data are processed and loaded into the EDW but are not included in any registry aggregate computations until corrected. Facilities are notified about data submission problems and provided an opportunity to resubmit data. Finally, a "green light" means that a submission has passed all integrity and quality checks. Such submissions are loaded to the EDW. After passing the DQR, data are loaded into a common EDW that houses data from all registries and included for all registry aggregate computations. In a secondary transaction process, data are loaded into registry-specific, dimensionally modeled data marts.

There is no sampling of patient data allowed within the contractual terms of participation in the CARE Registry in NCDR. The registry is designed to include 100 percent of consecutive adult patients who undergo a carotid revascularization procedure at participating institutions. Section 2.b of the NCDR Master Agreement with participants includes 'Participant Responsibilities': "b. Use of ACCF Data Set and ACCF-Approved Software. Participant will submit a data record on each patient who receives medical care and who is eligible for inclusion in the Registries in which Participant is participating under this Agreement." Adult patients, ages 18 years and older, who have an carotid revascularization procedure. Patients are selected for inclusion by reviewing existing medical records and no direct interaction with the patient will be required outside of the normal course of care. There will be no discrimination or bias with respect to inclusion on the basis of sex, race, or religion.
Patient confidentiality is preserved as the data are in aggregate form. The Care Registry dataset, comprised of approximately 250, data elements was created by a panel of experts using available ACC-AHA guidelines, data elements and definitions, and other evidentiary sources. Private health information (PHI), such as social security number, is collected. The intent for collection of PHI is to allow for registry interoperability and the potential for future generation of patient-level drill downs in Quality and Outcomes Reports. Registry sites can opt out of transmitting direct identifiers to the NCDR, however, so inclusion of direct identifiers in the registry is at the discretion of the registry participants themselves. When using the NCDR web-based data collection tool, direct identifiers are entered but a partition between the data collection process and the data warehouse maintains the direct identifiers separate from the analysis datasets. The minimum level of PHI transmitted to the ACCF when a participant opts out of submitting direct identifiers meets the definition of a Limited Dataset as such term is defined by the Health Insurance Portability and Accountability Act of 1996.
Data collection within the NCDR conforms to laws regarding protected health information. Patient confidentiality is of utmost concern with all metrics. The proposed measure does not include a patient survey. Physician and/or institutional confidentiality CARE Registry. No testing, time, risk, or procedures beyond those required for routine care will be imposed. The primary risk associated with this measure is the potential for a breach of patient confidentiality. The ACCF has established a robust plan for ensuring appropriate and commercially reasonable physical, technical, and administrative safeguards are in place to mitigate such risks.

Data are maintained on secure servers with appropriate safeguards in place. The project team periodically reviews all activities involving protected health information to ensure that such safeguards including standard operating procedures are being followed.

The procedure for notifying the ACCF of any breach of confidentiality and immediate mitigation standards that need to be followed is communicated to participants. ACCF limits access to Protected Health Information, and to equipment, systems, and networks that contain, transmit, process or store Protected Health Information, to employees who need to access the PHI for purposes of performing ACCF's obligations to participants who are in a contractual relationship with the ACCF. All PHI are stored in a secure facility or secure area within ACCF's facilities which has separate physical controls to limit access, such as locks or physical tokens. The secured areas are monitored 24 hours per day, 7 days per week, either by employees or agents of ACCF by video surveillance, or by intrusion detection systems.

Each participant who has access to the NCDR website must have a unique identifier. The password protected webpages have implemented inactivity time-outs. Encryption of wireless network data transmission and authentication of wireless devices containing NCDR Participant's information ACCF's network is required. Protected Health Information may only be transmitted off of ACCF's premises to approved parties, which shall mean: A subcontractor who has agreed to be bound by the terms of the Business Associate Agreement between the ACCF and the NCDR Participant.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
The ACCF's program the National Cardiovascular Data Registry (NCDR) provides evidence based solutions for cardiologists and other medical professionals committed to excellence in cardiovascular care. NCDR hospital participants receive confidential benchmark reports that include access to measure macro specifications and micro specifications, the eligible patient population, exclusions, and model variables (when applicable). In addition to hospital sites, NCDR Analytic and Reporting Services provides consenting hospitals' aggregated data reports to interested federal and state regulatory agencies, multi-system provider groups, third-party payers, and other organizations that have an identified quality improvement initiative that supports NCDR-participating facilities. Lastly, the ACCF also allows for licensing of the measure specifications outside of the Registry. For calendar year 2014 the annual pricing for hospitals, NCDR Analytic and Reporting Services, and licensing of measure specifications ranges from $2900-$50,000.
Measures that are aggregated by ACCF and submitted to NQF are intended for public reporting and therefore there is no charge for a standard export package. However, on a case by case basis, requests for modifications to the standard export package will be available for a separate charge.

There is no added procedural risk to patients through their hospital's involvement in the CARE Registry. No testing, time, risk, or procedures beyond those required for routine care will be imposed.

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.

Planned	Current Use (for current use provide URL)
Public Reporting	Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

	CARE Registry https://www.ncdr.com/webncdr/care/

4a.1. **For each CURRENT use, checked above, provide:**

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

CARE Registry of the National Cardiovascular Data Registry of the American College of Cardiology

4a.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?*)
Plan is to publically report in the future.

4a.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.*)
ACC is committed to implementing this measure. ACC is an authorized organization to receive CMS data through the ResDAC application process. Unfortunately, it has been determined by ResDAC that this authorization does not permit use of CMS for performance measure reporting purposes, either to hospitals or for public display. ACC is currently in process of applying to be a Qualified Entity. It is unclear if this pathway will permit measure implementation. ACC also is commenting on and tracking proposed language in 21st Century Cures legislation, which does appear to create a pathway for use of CMS data for this type of reporting purpose.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

Not available, initial endorsement

4b.2. **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.**

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended consequences identified.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

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

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

5a. Harmonization

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 completely harmonized?
Yes

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

5b. Competing Measures

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

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
No competing measures.

Appendix

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

Attachment **Attachment:** CARE_v109_CAS_DataDictionaryDefinitionsOnly-635707384238518946.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American College of Cardiology
Co.2 **Point of Contact:** Penelope, Solis, comment@acc.org, 202-375-6576-
Co.3 **Measure Developer if different from Measure Steward:**
Co.4 **Point of Contact:**

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.
SQOC—Leadership committee that oversaw broad issues and approved submission of given metric to NQF.
Fred Masoudi, David Malenka, Thomas Tsai, Matt Reynolds, David Shahian, John Windle, Fred Resnic, John Moore, Deepak Bhatt, James Tcheng, Jeptha Curtis, Paul Chan, Matt Roe, John Rumsfeld
Clinical SubWorkgroup-oversaw NQF application components
Jeptha Curtis-chair
Christopher White, Thomas Tsai, John Rumsfeld, Fred Masoudi
CARE/PVI Transition Workgroup -Provides strategic direction for the Registry and monitors research and clinical activities.
Chris White, Kalon Ho, Ken Rosenfield, Bobby Yeh, Michael Jaff, Thomas Tsai, P. Michael Grossman, Herb Aronow, H. Vernon Anderson

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?**
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

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 #: 2712
Measure Title: Statin Use in Persons with Diabetes
Measure Steward: Pharmacy Quality Alliance (PQA, Inc.)
Brief Description of Measure: The percentage of patients ages 40 – 75 years who were dispensed a medication for diabetes that receive a statin medication.
Developer Rationale: The American College of Cardiology/American Heart Association (ACC/AHA) guidelines recommend moderate- to high-intensity statin therapy for primary prevention for persons aged 40-75 years with diabetes (class I recommendation). Guideline: 2013 ACC/AHA Guideline on Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: a Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. (1)

The measure reflects this new clinical guideline and will promote appropriate treatment of patients with diabetes (age 40-75) to reduce their risk of cardiovascular disease and complications.

Prescription claims data are used as a proxy for diabetes diagnosis in this measure as well as other PQA and HEDIS measures. Medical data used in testing confirmed that the denominator criteria of two prescription claims for a hypoglycemic agent identified a population where a great majority had a diagnosis of diabetes during the measurement year. These criteria also included very few persons with select conditions (i.e., polycystic ovarian syndrome, gestational diabetes or diabetes secondary to another condition) that were considered for exclusion from the measure.
This measure uses only prescription claims as a source of data resulting in the inability to identify individuals with contraindications to statin therapy or other medical exceptions. Therefore the performance rate goal for this measure is not intended to reach 100%.

1. Stone NJ, Robinson J, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero ST, Smith SC Jr, Watson K, Wilson PWF. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2013;00:000–000. Accessed 2/3/2014 http://circ.ahajournals.org/content/early/2013/11/11/01.cir.0000437738.63853.7a.full.pdf

Numerator Statement: The number of patients in the denominator who received a prescription fill for a statin or statin combination during the measurement year.
Denominator Statement: The denominator includes subjects aged 41 years – 75 years as of the last day of the measurement year who are continuously enrolled during the measurement period. Subjects include patients who were dispensed two or more prescription fills for a hypoglycemic agent during the measurement year.
Denominator Exclusions: Patients in Hospice (Medicare Part D) are excluded from this measure. Medicare prescription claims for persons in hospice are not covered by Part D.

Measure Type: Process
Data Source: Administrative claims
Level of Analysis: Health Plan, Population : National
Is this an eMeasure? ☐ Yes ☒ No **If Yes, was it re-specified from a previously endorsed measure?** ☐ Yes ☐ No

☐ **Is this a MAINTENANCE measure submission?** ☐ Yes ☒ **No, this is a NEW measure submission.**
For a MAINTENANCE, what is the Original Endorsement Date: N/A **Most Recent Endorsement Date: N/A**

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for an *process* 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

The developer provides the following evidence for this process measure:

- This is a health plan-level process that calculates the percentage of patients ages 40 – 75 years who were dispensed a medication for diabetes that receive a statin medication.
- The developer provides the path from initiating statin therapy in patients with diabetes to a reduction in the risk of atherosclerotic cardiovascular disease.
- The measure is based on an ACC/AHA guideline for the primary prevention in individuals with diabetes. The evidence for this guideline was graded Level A, Class 1.
- Developers provided a summary of the quantity, quality, and consistency of the evidence.

Questions for the Committee:

- *Is the evidence directly applicable to the process being measured?*
- *Is the process proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- The developer provided 2012 and 2013 data for the following health plans, though individual descriptive statistics are not provided by year and insurance type for 2012:

	Type of plan	# of patients(N)	Min	Max	Mean	SD	IQR
2012	Medicare	1,807,725	59.1%	67.6%	62.8%	6.6%	-
	Commercial	16,615,029					
	Medicaid	665,715					
2013	Medicare Part D	23,185,246	66.1%	100%	72.5%	8.3%	6.2%

- The developer provided additional information on the Medicare Drug Benefit: The Medicare Drug Benefit is provided by private prescription drug plans (PDPs) that offer drug-only coverage, or through Medicare Advantage health plans that offer both prescription drug and health care coverage (known as MA-PDs). Currently there are 23.9 million lives in PDPs and 15.3 million lives in MA-PDs:
  - Medicare PDP – 61%
  - Medicare MA-PDP – 39%
- The disparities data provided by the developer includes the rates for the Low Income Subsidy (LIS) population; disparities data from the literature was not provided.
- The developer provided 2012 data for 3 insurance types and 2015 (Jan-Mar) data for eight Part D contracts for the Low Income Subsidy (LIS) and non-LIS population:

	Type of plan	Rate

2012	Medicare	73.6%
	Commercial	60.4%
	Medicaid	59.1%
2015	Medicare Part D: LIS	67.3%
	Medicare Part D: Non-Lis	67.1%

- The developer states LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. Medicare beneficiaries apply for the LIS with the Social Security Administration or their State Medicaid agency. The following groups automatically qualify for LIS without applying:
  - Full–benefit dual eligible (Medicare and Medicaid eligible)
  - Supplemental Security Income recipients
  - Medicare beneficiaries who participate in the Medicare Saving Programs
  - QMB – qualified Medicare beneficiary
  - SLMB – specified low-income Medicare beneficiary
  - QI – qualifying individual

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Is the SC aware of additional evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - There has been a significant body of literature over the years to support the concept that people with diabetes have better outcomes with lower cholesterol; and that statins appear to have additional protective properties beyond that of cholesterol reduction alone.
   - Health Outcome
     - Intermediate clinical outcome
     - Process
     - Structure
     - Efficiency

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- This process identifies persons with diabetes via prescriptions for a hypoglycemic agent and then identifies subsequent statin use. The use of a statin in this population may lead to a reduction of risk in cardiovascular complications.

1b. ***Committee's Comments on Performance Gap:***

- Yes. While there has been a much greater recognition of the value of statin therapy among specialists and PCPs, the adherence has not been at the level that is appropriate. Much of this gap may have come from the over concern for statin induced myopathy so that patients and doctors have had an over concern about the use of statins with diabetes.
- Not provided. This is a missed opportunity because many minority women have higher risks for diabetes, thus cardiovascular disease.

1c. ***Committee's Comments on Composite Performance Measure:***

- Not a composite measure
- The NQF Measure 0729 is a Related measure. It is Optimal Diabetes Care - Cholesterol Statin Use Component. The Measure Steward is MN Community Measurement.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure assess the percentage of patients ages 40 – 75 years diagnosed with diabetes who were dispensed a statin medication. To be included in the denominator, patients must be 41 years by the end of the measurement year.
- The measure uses prescription administrative claims data which is available electronically. The eligible medications used to calculate the numerator/denominator are provided.
- The measure uses diabetic medications as proxy information for a diagnosis of diabetes. As the measure exclusively uses prescription administrative claims data (i.e., clinical data is not collected) to calculate performance, this measure does *not* assess diabetic patients who are treated non-pharmaceutically, patients with a contradiction to statins (e.g., allergy, intolerance, refusal), diabetic patients with cholesterol monitored and within normal limits, nor patients who are exclusively taking over-the-counter cholesterol-lowering medications.
- Patients with gestational diabetes, steroid-induced diabetes, and poly-cystic ovarian disease may be prescribed hypoglycemic agents and are not excluded from the measure.
- The measure description states "a" single medication identifies patients as diabetic, while rationale and the calculation algorithm state "two prescription claims for a hypoglycemic agent identified a population where a great majority had a diagnosis of diabetes".
- Those persons receiving hospice care at any point during the measurement year are excluded, though the developer notes a limitation to this exclusion because hospital enrollment data may not be routinely available to non-Medicare plans, such as Medicaid and Commercial lines of business.
- The measure is stratified by insurance type: Commercial, Medicaid, and Medicare, and it is not risk adjusted.
- A clear calculation algorithm for the measure is provided. The measure is not risk adjusted.

Questions for the Committee:

- *Should all hospice patients be excluded, including those from non-Medicare plans?*
- *Should all patients prescribed hypoglycemic agents be prescribed statins?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- The developer used signal to noise analysis with a mixed effect logistic regression model to examine the variability in performance measure score at the health plan level, which is the unit of analysis specified for the prescription-only claims measure. The developer used a signal to noise methodology other than The Reliability of Provider Profiling (2009) familiar to the Committee. The developer states this type of reliability testing is appropriate for this type of measure, and that the testing differentiates the true difference between measured entities (the signal) to random measurement error (the noise). The developer tested (and rejected) the null hypothesis that performance does not vary across the units being measured. From their findings, observed performance variation is at least partly due to true differences (i.e. signal) and is not entirely due to random statistical variation (i.e. noise), though the developer should provide further

interpretation and quantification of the results. The developer is encouraged to provide clarification for reliability testing methods and results.

- A likelihood-ratio (LR) test was also performed to determine if a model with random effects would fit the data better than a standard logistic regression model without random effects. This test was conducted with the Medicare Part D Prescription Drug Event (PDE) data described in sections 1.2- 1.7. It does not include Medicare Advantage, Medicaid & Commercial insurance data. The developer is encouraged to provide clarification for reliability testing methods and results.

Table 1. Mixed Effect Logistic Regression Model, Statin Use in Persons with Diabetes measure rate comparison across Part D plans

	Coefficient	Standard Error	Z	p-value
Intercept	1.018	0.009	110.56	<0.001
	Estimate	**Standard Error**	**95% Confidence Interval**	
Random Effects	0.224	0.008	0.210	0.239

- The developer states these results indicate that there are significant differences in performance measure scores between plans, which allows for discrimination between high performing plans and low performing plans. The developer reports that based on these results, the measure is considered to be reliable.
- The testing methodology and results are pending review.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The clinical practice guideline supporting this measure recommends the use of statin therapy in persons with diabetes 40 to 75 years of age.

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Do the evidence state patients with gestational diabetes, steroid-induced diabetes, and poly-cystic ovarian disease should be prescribed statins?*

2b2. Validity testing

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.

- Face Validity - The developers provide a 7 step consensus based measure development and testing process, and state that 34 of the 38 (89.5%) members of the PQA Workgroup who developed the SUPD measure, agreed that the measure could differentiate the quality of care.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

○ *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Those persons receiving hospice care at any point during the measurement year are excluded. No testing was performed on this exclusion, as the data source, prescription claims data, do not contain claims for persons that are in hospice care.
- The developer notes a limitation to this exclusion because hospital enrollment data may not be routinely available to non-Medicare plans such as Medicaid and Commercial lines of business.

Questions for the Committee:

- *Should all hospice patients be excluded, including those from non-Medicare plans?*
- *Should patients with gestational diabetes, Steroid-induced diabetes, and poly-cystic ovarian disease be excluded from the measure?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This process measure is not risk adjusted.

2b5. Meaningful difference:

- The overall mean performance on this measure varied between health plans: 59.1% (Medicaid) to 73.6% (Medicare) in 2012. The rates for 2013 also showed significant variation for Medicare Part D from 66.07% to 100% (SD = 8.31%).

Question for the Committee:

○ *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- This is not appropriate as the measure is not specified for multiple data source.

2b7. Missing Data

- All data elements required to calculate the measure are available in the prescription claims; no missing data was found.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- The concern that I have is in the definition of diabetes used in this measure as being medication managed. This means those individuals with diabetes not on medications are not included in the numerator or denominator. This is different than the determinations of diabetes used by NCQA where at least two separate claims with the ICD codes for diabetes may be used for the inclusion of people with diabetes. As a result this measure is somewhat distorted.
- Reliability testing 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 procise.

2a2.: ***Committee's Comments on Reliability-Testing:***

- The measure has been used sufficiently to have reliability for the measured population
- Reliable data points used, populations included increased substantially over time.

2b1.: ***Committee's Comments on Validity-Specifications***:

- As in the section above. The limitation of people on pharmacologic therapy for diabetes means that these are likely to be people with greater degrees of glucose intolerance, and are more likely to be people who would be medication compliant. So the measure is valid for its defined population but not valid for all diabetic people.
- The Validity testing demonstrates that the measure data elements are correct and/or the measure scores correctly reflects the quality of care provided adequately identifying differences in quality.

2b2.: ***Committee's Comments on Validity-Testing:***

- This has been considered a standard and valid measure---but only for the population of people with diabetes who are on mediation management.
- Critical data elements used in readability testing.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- The reliability and validity of prescription claims data tested and each one evaluate 4d in the literature and deemed reliable and valid for this measure.
- As above. The risk adjustment of people not on medication for their diabetes means that the population being measured is different than the population of all people with diabetes. This may mean that if the population included all diabetics the results may be different. Confounding factors for non-prescription of medication may include:
  1. Severity of glucose intolerance.
  2. Medication adherence.
  3. Social determinant such as access to care, numeracy and literacy

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- The data source is electronically abstracted from administrative claims and readily available from health plan prescription claims data and enrollment data.
- Health plans already obtain prescription claims data for payment therefore there is no extra burden/cost in the collection of data for this measure.
- The developer states organizations must obtain permission to use this measure and they may require a license, though no details or costs are provided.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- This has been an existing measure so is quite feasible
- Data Elements generated as byproduct of care process.

Criterion 4: Usability and Use

4. Usability and 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.

- The measure is currently reported by CMS to all Medicare Part D health plan sponsors in the monthly Patient Safety Reports for quality improvement.
- URAC is planning to add this measure as an exploratory measure to their accreditation programs for Community Pharmacy, PBM, and Mail Service.
- CMS is considering this measure as a new 2017 display measure (using 2015 data) and as a possible 2018 Star Rating measure (using 2016 data).
- The developers state no unintended consequences were identified during testing of this measure.

***Questions for the Committee**:*

- *Is the measure publicly reported?*
- *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?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- As an accountability measure, it has been applied to clinical practices, integrated delivery systems and to health plans--all of which have some ability to impact the results through their actions.
- Expected to be used in at least one accountability application within 3 years and publicly reported in 6 years.

Criterion 5: Related and Competing Measures

- 0729: Optimal Diabetes Care- Cholesterol Statin Use Component
- The developer states that the measure specifications are not harmonized because 2712 is a health plan measure that addresses only appropriate use of statins in diabetics age 40-75 while 0729 is a clinician level composite measure that addresses various aspects of care for patients with IVD.

Pre-meeting public and member comments

Comment by: Ashish R. Trivedi, Pharm.D.
Organization: SPI-Lilly

Comment#5113: Lilly is supportive of the direction of the new guidelines focused on treating and reducing cardiovascular risk (vs treating to LDL-C targets) in patients with diabetes, who represent a large population of patients at substantially increased risk for ASCVD (atherosclerosis cardiovascular disease) events [Stone et al, 2013]. However, we would like to point out that comprehensive and routine lipoprotein lipid assessment is still an integral part of managing risk in patients with ASCVD [Jacobson et al, 2015]. In addition, clinical trial data indicates significant residual cardiovascular risk in ASCVD patients treated with statins, even in the setting of optimal LDL-C reduction (eg, <70 mg/dL and <100 mg/dL), thus highlighting the need to consider alternative CV risk reduction algorithms beyond the focus on LDL-C levels and/or the use of statins [Cannon et al 2004, LaRosa et al 2005, Pedersen et al 2005].

References

- Stone NJ, Robinson JG, Lichtenstein AH, et al. ACC/AHA Prevention Guideline: 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of

the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129:25 suppl 2 S1-S45, doi:10.1161/01.cir.0000437738.63853.7a
- Jacobson TA, Ito MK, Maki KC, et al. National Lipid Association recommendations for patient-centered management of dyslipidemia: part 1 – Full Report. J Clin Lipidol. 2015; 9(2), 129–169. DOI: http://dx.doi.org/10.1016/j.jacl.2015.02.003
- Cannon CP, Braunwald E, McCabe CH, et al. Intensive versus moderate lipid lowering with statin after acute coronary syndromes. N Engl J Med. 2004; 350:1495–1504.
- LaRosa JC, Grundy SM, Waters DD, et al. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. N Engl J Med. 2005; 352:1425–1435.
- Pedersen TR, Faergeman O, Kastelein JJ, et al. High-dose atorvastatin vs usual-dose simvastatin for secondary prevention after myocardial infarction: the IDEAL study: a randomized controlled trial. JAMA. 2005; 294:2437–2445.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Statin Use in Persons with Diabetes

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: 6/29/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: The percentage of patients ages 40 – 75 years who were dispensed a medication for diabetes that receive a statin medication.

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

For people aged 40-75, there are considerable potential benefits from initiating stain therapy, especially for those with a cardiovascular risk factor such as diabetes. This process measure identifies persons with diabetes via prescriptions for a hypoglycemic agent and then identifies subsequent statin use. The use of a statin in this population may lead to a reduction of risk in cardiovascular complications.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Stone NJ, Robinson J, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero ST, Smith SC Jr, Watson K, Wilson PWF. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation.* 2013;00:000–000. Accessed 6/2/2015
http://circ.ahajournals.org/content/early/2013/11/11/01.cir.0000437738.63853.7a.full.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Page 31 of the guidelines state:

"4.5. Primary Prevention in Individuals with Diabetes**:** A high level of evidence supports the use of moderate-intensity statin therapy in persons with diabetes 40 to 75 years of age. The only trial of high-intensity statin therapy in primary prevention was performed in a population without diabetes. However, a high level of evidence was considered for event with statin therapy reduction in individuals with a ≥7.5% estimated 10-year ASCVD risk (Section 4.6) who did not have diabetes to recommend high-intensity statin therapy preferentially for individuals with diabetes and a ≥7.5% estimated 10-year ASCVD risk (Section 4.7). This consideration for those with diabetes 40 to 75 years of age recognizes that these

individuals are at substantially increased lifetime risk for ASCVD events and death. Moreover, individuals with diabetes experience greater morbidity and worse survival following the onset of clinical ASCVD.

In persons with diabetes <40 or >75 years of age, statin therapy should be individualized based on considerations of ASCVD risk reduction benefits, the potential for adverse effects and drug-drug interactions, and patient preferences (Figure 4)."

ASCVD = atherosclerotic cardiovascular disease

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

The National Heart, Lung, and Blood Institute (NHLBI) Grading methodology was used for this recommendation. This guideline is graded A, which is defined as: Strong recommendation - There is high certainty based on evidence that the net benefit is substantial.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.

(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

The following table contains the grades and associated definitions for grading system recommendations:

Grade	Strength of Recommendation*
A	**Strong recommendation** There is high certainty based on evidence that the net benefit† is substantial.
B	**Moderate recommendation** There is moderate certainty based on evidence that the net benefit is moderate to substantial, or there is high certainty that the net benefit is moderate.
C	**Weak recommendation** There is at least moderate certainty based on evidence that there is a small net benefit.
D	**Recommendation against** There is at least moderate certainty based on evidence that it has no net benefit or that risks/harms outweigh benefits.
E	**Expert opinion ("There is insufficient evidence or evidence is unclear or conflicting, but this is what the Work Group recommends.")** Net benefit is unclear. Balance of benefits and harms cannot be determined because of no evidence, insufficient evidence, unclear evidence, or conflicting evidence, but the Work Group thought it was important to provide clinical guidance and make a recommendation. Further research is recommended in this area.

N	**No recommendation for or against ("There is insufficient evidence or evidence is unclear or conflicting.")** Net benefit is unclear. Balance of benefits and harms cannot be determined because of no evidence, insufficient evidence, unclear evidence, or conflicting evidence, and the Work Group thought no recommendation should be made. Further research is recommended in this area.

*In most cases, the strength of the recommendation should be closely aligned with the quality of the evidence; however, under some circumstances, there may be valid reasons for making recommendations that are not closely aligned with the quality of the evidence (e.g., strong recommendation when the evidence quality is moderate, like smoking cessation to reduce CVD risk or ordering an ECG as part of the initial diagnostic work-up for a patient presenting with possible MI). Those situations should be limited and the rationale explained clearly by the Work Group.

†Net benefit is defined as benefits minus risks/harms of the service/intervention

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

See citation in 1a.4.1; pages 6-8

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

The treatment addressed in the evidence review is moderate-intensity statin therapy for adults 40 to 75 years of age with diabetes mellitus.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

The quality of the quoted evidence is graded using the American College of Cardiology/American Heart Association (ACC/AHA) Level of Evidence (LOE) grading system. The evidence was graded a Level A, Class 1.

The estimate of the precision of the treatment effect is graded Level A. This is defined as: Multiple (3-5) population risk strata evaluated; general consistency of direction and magnitude of effect.

The size of the treatment effect is graded Class 1. This is defined as: Benefit >>> Risk. No additional studies are needed. Procedure/treatment should be performed/administered.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

The following table contains the grades and associated definitions for grading the strength of the evidence:

Grade	Estimate of the precision of the treatment effect

A	Multiple (3-5) population risk strata evaluated; general consistency of direction and magnitude of effect.
B	Limited (2-3) population risk strata evaluated.
C	Very limited (1-2) population risk strata evaluated.
Grade	**Size of treatment effect**
Class I	Benefit >>> Risk. No additional studies are needed. Procedure/treatment should be performed/administered.
Class IIa	Benefit >> Risk. Additional studies with focused objectives are needed. It is reasonable to perform procedure/administer treatment.
Class IIb	Benefit ≥ Risk. Additional studies with broad objectives are needed; additional registry data would be helpful. It is not unreasonable to perform procedure/administer treatment.
Class III	Risk ≥ Benefit. No additional studies are needed. Procedure/treatment should not be performed/administered since it is not helpful and may be harmful.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 1998-2010

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Five (5) studies were included in the body of evidence for this recommendation. Three (3) studies are randomized control trials (RCT). Two studies are meta-analyses of RCTs, one that included 26 trials and one that included 14 trials.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

The overall quality of the evidence across studies is graded using the ACC/AHA Level of Evidence (LOE) grading system. The evidence was graded a Level A, Class 1.

The estimate of the precision of the treatment effect is graded Level A. This is defined as: Multiple (3-5) population risk strata evaluated; general consistency of direction and magnitude of effect.

The size of the treatment effect is graded Class 1. This is defined as: Benefit >>> Risk. No additional studies are needed. Procedure/treatment should be performed/administered.

The populations included in the studies included the intended target population of this measure, persons with diabetes 40 to 75 years of age. The quality of the evidence, estimate of the precision of the treatment effect and size of the treatment effect each received the highest grading, providing a high degree of confidence.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

In the RCTs reviewed, initiation of moderate to high intensity therapy is a critical factor in reducing arteriosclerotic cardiovascular disease (ASCVD) events.

The review of the RCTs showed that statin use reduced the incidence of major coronary and vascular events, with a relative risk (RR) range of 0.63 to 0.82 (*P* range from 0.003 to 0.001). Statin use also reduced the mortality rate, with a RR range from 0.63 to 0.90 (*P* range from 0.059 to 0.001).

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Adverse events were studied across the RCTs that were reviewed. One meta-analysis found that adverse events were generally mild, but 17 RCTs reported on increased risk of development of incident diabetes [Odds Ratio (OR) 1.09; 95% CI 1.02–1.17, $P = 0.001$, $I^2 = 11\%$]. One study found that there was no significant difference in the incidence of malignant neoplasms or other serious adverse events, while another found that there was no evidence that reduction of LDL cholesterol with a statin increased cancer incidence (RR per 1.0 mmol/L LDL cholesterol reduction 1.00, 95% CI 0.96–1.04), cancer mortality (RR 0.99, 95% CI 0.93–1.06), or other non-vascular mortality. Across studies, the findings were consistent with the findings that that the benefit of stain therapy greatly exceeds any known risks.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

NA

1a.8 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.8.1 What process was used to identify the evidence?

NA

1a.8.2. Provide the citation and summary for each piece of evidence.

NA

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
SUPD_Template_MeasSubm_Evidence_FINAL_062615.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
The American College of Cardiology/American Heart Association (ACC/AHA) guidelines recommend moderate- to high-intensity statin therapy for primary prevention for persons aged 40-75 years with diabetes (class I recommendation).
Guideline: 2013 ACC/AHA Guideline on Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: a Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. (1)

The measure reflects this new clinical guideline and will promote appropriate treatment of patients with diabetes (age 40-75) to reduce their risk of cardiovascular disease and complications.

Prescription claims data are used as a proxy for diabetes diagnosis in this measure as well as other PQA and HEDIS measures. Medical data used in testing confirmed that the denominator criteria of two prescription claims for a hypoglycemic agent identified a population where a great majority had a diagnosis of diabetes during the measurement year. These criteria also included very few persons with select conditions (i.e., polycystic ovarian syndrome, gestational diabetes or diabetes secondary to another condition) that were considered for exclusion from the measure.
This measure uses only prescription claims as a source of data resulting in the inability to identify individuals with contraindications to statin therapy or other medical exceptions. Therefore the performance rate goal for this measure is not intended to reach 100%.

1. Stone NJ, Robinson J, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero ST, Smith SC Jr, Watson K, Wilson PWF. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2013;00:000–000. Accessed 2/3/2014 http://circ.ahajournals.org/content/early/2013/11/11/01.cir.0000437738.63853.7a.full.pdf

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

This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.
Testing results for this measure was performed on data for calendar years 2012 and 2013.

For the 2012 data, results were calculated for one Medicare plan (N=1,807,725), one Commercial plan (N=16,615,029) and one Medicaid plan (N=665,715). The measure rates ranged from 59.1% to 67.6%, with a mean of 62.8% and a standard deviation of 6.6%.

For the 2013 data, results were calculated for 736 Medicare Part D plans (N=23,185,246). The measure rates range from 66.1% to 100%, with a mean of 72.5% and a standard deviation of 8.3%. The Interquartile Range (IQR) is 6.2%.

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

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
For calendar year 2012, the measure rates were calculated by three different insurance types. The rate for the Commercial insurance population is 60.4%. The rate for the Medicare population is 73.6%, and the rate for the Medicaid population is 59.1%.

Data from January-March 2015 for eight (8) Part D contracts show that the measure rate for the Low Income Subsidy (LIS) population is 67.3%, while the rate in the Non-LIS population is 67.1%.

Definition: Medicare Low Income Subsidy (LIS)
A subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. Medicare beneficiaries apply for the LIS with the Social Security Administration or their State Medicaid agency.

The following groups automatically qualify for LIS without applying:
- Full–benefit dual eligible's (Medicare and Medicaid eligible)
- Supplemental Security Income recipients
- Medicare beneficiaries who participate in the Medicare Saving Programs
  - QMB – qualified Medicare beneficiary
  - SLMB – specified low-income Medicare beneficiary
  - QI – qualifying individual

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)
The measure addresses:
- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
This measure addresses two highly prevalent diseases, diabetes and cardiovascular disease (CVD). There are 29 million people in the

United States with diabetes, 9.3% of the population. (1) Cardiovascular disease is the leading cause of death in the United States.(2) Diabetes is a significant risk factor for CVD and diabetes can be viewed as a high-risk state for CVD, similar to having coronary heart disease.(3) The ACC/AHA Guidelines indicate that persons with diabetes who are 40-75 years of age are at increased risk of developing atherosclerotic cardiovascular disease events and statin therapy is used to decrease this risk. (4)

PQA testing data from CMS and a commercial database also demonstrates that the measure addresses a large number of people. In a commercial population of around 16 million, there were 1.3 million people meeting the measure denominator. Nearly 5 million Medicare beneficiaries met the denominator criteria of a total population of 23 million.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. American Diabetes Association. Statistic about Diabetes Data from the 2012 National Diabetes Fact Sheet. Accessed 6/7/2015 at http://www.diabetes.org/diabetes- basics/statistics

2. 2013 Mortality Multiple Cause Micro-data Files. Detailed Tables for the National Vital Statistics Report (NVSR) "Deaths: Final Data for 2013."
http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_02.pdf

3. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III): Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. 106:3143–3421, 2002

4. Stone NJ, Robinson J, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero ST, Smith SC Jr, Watson K, Wilson PWF. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2013;00:000–000.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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 : Hyperlipidemia, Endocrine : Diabetes

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Health and Functional Status

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.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool

(MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The number of patients in the denominator who received a prescription fill for a statin or statin combination during the measurement year.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The measurement period is generally a 12 month calendar year and extends through the last day of the enrollment period or until death or disenrollment.

S.6. **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, 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.
The number of patients in the denominator who received a prescription fill for a statin or statin combination during the measurement year. Statin medications for this measure include: lovastatin, rosuvastatin, fluvastatin, atorvastatin, pravastatin, pitavastatin, simvastatin. Statin combination medications for this measure include: niacin & lovastatin, atorvastatin & amlodipine, niacin & simvastatin, sitagliptin & simvastatin, ezetimibe & simvastatin, ezetimibe & atorvastatin. Note: The active ingredients are limited to oral formulations only.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator includes subjects aged 41 years – 75 years as of the last day of the measurement year who are continuously enrolled during the measurement period. Subjects include patients who were dispensed two or more prescription fills for a hypoglycemic agent during the measurement year.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Subjects are included if they are age 41-75 at the end of the measurement year. Subjects should be continuously enrolled during the measurement period. To determine continuous enrollment using enrollment data, for a Medicaid beneficiary for whom enrollment is verified monthly, the member may not have more than a 1-month gap in coverage (i.e., a member whose coverage lapses for 2 months [60 consecutive days] is not considered continuously enrolled). Subjects are included in the denominator if they were dispensed two or more prescription fills for a hypoglycemic agent during the measurement year. Hypoglycemic medications for this measure include:

Biguanides and Biguanide Combination Products: Metformin, pioglitazone & metformin, rosiglitazone & metformin, repaglinide & metformin, sitagliptin & metformin IR & SR, saxagliptin & metformin SR, linagliptin & metformin, glyburide & metformin, glipizide & metformin, alogliptin & metformin

Sulfonylureas and Sulfonylurea Combination Products: chlorpropamide, glipizide & metformin, glimepiride, glipizide, glyburide & metformin, glyburide, rosiglitazone & glimepiride, pioglitazone & glimepiride, tolazamide, tolbutamide

Meglitinides and Meglitinide Combination Products: nateglinide, repaglinide, repaglinide & metformin

Alpha- Glucosidase Inhibitors: acarbose, miglitol

Thiazolidinediones and Thiazolidinedione Combination Products: pioglitazone, pioglitazone & glimepiride, pioglitazone & metformin, rosiglitazone, rosiglitazone & glimepiride, rosiglitazone & metformin, alogliptin & pioglitazone

Incretin Mimetic Agents: exenatide, dulaglutide, liraglutide, albiglutide

Amylin Analogs: pramlintide

DPP-IV Inhibitors and DPP-IV Inhibitor Combination Products: sitagliptin, linagliptin, alogliptin, saxagliptin, alogliptin & metformin, alogliptin & pioglitazone, linagliptin & metformin, sitagliptin & metformin IR & SR, saxagliptin & metformin SR, sitagliptin & simvastatin

Insulins: insulin aspart, insulin aspart Protamine & Aspart, insulin detemir, insulin glargine, insulin glulisine, insulin isophane & regular human insulin, insulin isophane (human N), insulin lispro, insulin lispro Protamine & Insulin lispro, insulin regular (human R), insulin regular (human) inhalation powder

Sodium glucose co-transporter2 (SGLT2) Inhibitors: canagliflozin, dapagliflozin, emapaglifozin

Note: Excludes nutritional supplement/dietary management combination products.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Those persons receiving hospice care at any point during the measurement year.
The exclusion uses enrollment data.

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Hospice status may not be identifiable in all non-Medicare prescription drug plan benefits. The exclusion will be for any person receiving hospice care during the measurement year. It is a limitation of the measure if enrollment data for line of business (e.g. Medicaid, Commercial) does not routinely include this information and therefore cannot use the exclusion in the measure calculation.
Limitation: Hospice enrollment data may not be routinely available to non-Medicare plans such as Medicaid and Commercial lines of business

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
This measure will be stratified by insurance product line. Rates for Commercial, Medicaid, and Medicare will be reported separately.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*

Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Denominator Calculation:
Step 1: Identify the eligible population that is 41-75 years of age as of the last day of the measurement period and that are continuously enrolled in the drug plan.
Step 2: Exclude any person that is in hospice (Medicare Part D)
Step 3: Identify those patients in Step 2 who were dispensed two or more prescription fills for a hypoglycemic agent during the measurement year.
The number of patients identified in Step 3 is the denominator for the measure.

Numerator Calculation:

Step 4: Of those patients identified in Step 3, identify the patients who received one or more prescription fills for a statin or statin combination during the measurement year.
The number of patients identified by completing Step 4 represents the numerator for this measure.

Step 5: Divide the numerator by the denominator and then multiply by 100 to obtain the rate (as a percentage) for the measure.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Health plan (e.g., Medicare, Medicaid, other) prescription claims data. Health Plan member enrollment information. This measure is intended to be reported by prescription drug plans that only have prescription claims and enrollment data.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Pharmacy
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
SUPD_measure_testing_attachment_FINAL_062615.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Statin Use in Persons with Diabetes
Date of Submission: 6/29/2015
Type of Measure:

☐Composite – ***STOP – use composite testing form***	☐Outcome (*including PRO-PM*)
☐Cost/resource	XProcess
☐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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource usemeasures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims andEHRs), section **2b6** also must be completed.
- Respond to all questions as instructedwith answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2.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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

2b4.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 sociodemographic 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.

2b5. 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.

2b6.If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data),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.

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 byaspect 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 forall 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.23***)	**Measure Tested with Data From:**
☐abstracted from paper record	☐abstracted from paper record
X administrative claims	X administrative claims
☐clinical database/registry	☐clinical database/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*).

For 2012, the data were derived from three data sources. The first is the Truven Health MarketScan® Commercial Claims and Encounters Database. These data represent the health services of approximately 139 million employees, dependents, and retirees in the United States with primary coverage through privately insured fee-for-service, point-of-service, or capitated health plans. There were more than 40 million lives in the database in for 2012.

The second was The MarketScan Medicare Supplemental and Coordination of Benefits Database (Medicare Supplemental Database). This database represents over 8.3 million lives and had more than 3.3 million covered lives in 2012. The database includes the Medicare-covered portion of payment (represented as Coordination of Benefits Amount, or COB), the employer-paid portion, and any out-of-pocket patient expenses. The data elements in this database are the same as those appearing in the Commercial Database, but pertain to patients with Medicare supplemental insurance.

The third data source is the MarketScan Medicaid Multi-State Database, which contains the medical, surgical and prescription drug experience of more than 19 million Medicaid enrollees from multiple states. It includes records of inpatient services, inpatient admissions, outpatient services, and prescription drug claims, as well as information about long term care and other medical care. Data on eligibility, service, and provider type are also included. There were more than 5 million lives in the database in 2012.

For 2013, the data used for testing came from ***three data sources***. For identification of prescription drugs, the Medicare Part D Prescription Drug Event (PDE) claims were used. To identify dates of birth and continuous enrollment, the Common Medicare Environment (CME) data source was used. To identify hospice enrollment, the Medicare Enrollment Database (EDB) was used.

1.3. What are the dates of the data used in testing? Two years of data were used for testing, from two different data sources. The years included calendar year 2012 and calendar year 2013.

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐group/practice	☐group/practice
☐hospital/facility/agency	☐hospital/facility/agency
X health plan	X health plan
☐other: Click here to describe	☐other: Click here to describe

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

For the calendar year 2012 data, the testing analysis included 25 health plans, including a national convenience sample of privately insured fee-for-service, point-of-service, capitated health plans, Medicare and Medicaid plans. The size and characteristics of the population are included at the patient level in 1.6.

For the calendar year 2013 data, the testing analysis included a convenience sample of 736 Medicare Part D prescription drug plans. The size and characteristics of the population are included at the patient level in 1.6.

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

For calendar year 2012, a total of 19,088,469 patients age 40-75 were included in the testing and analysis. This data can be stratified by insurance type, gender, and age. Of all patients, 16,615,029 (87.0%) are enrolled in a Commercial health insurance plan, 1,807,725 (9.5%) are enrolled in Medicare, and 665,715 (3.5%) are enrolled in Medicaid. Of all patients, 9,018,357 (47.2%) are male, and 10,070,112 (52.8%) are female. Patients by age group included 17,261,150 (90.4%) age 40-64 years, and 1,827,319 (9.6%) age 65-75.

For calendar year 2013, a total of 22,145,248 patients age 40-75 were included in the testing and analysis. This data can be stratified by gender and age. Of all patients, 9,943,095 (44.9%) are male, and 12,202,153 (55.1%) are female. Patients by age group included 5,743,719 (25.9%) age 40-64 years, and 16,401,529 (74.1%) age 65-75.

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.
N/A

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

No patient level or proxy sociodemographic status (SDS) variables were analyzed in the 2012 and 2013 testing data. We reviewed patient-level Low Income Subsidy (LIS) data from January-March 2015 for eight (8) Part D contracts, which showed that the measure rate for the LIS population is not significantly different from the rate in the Non-LIS population (67.3% vs 67.1%, respectively.)

2a2. RELIABILITY TESTING
***<u>Note</u>**: 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? (*maybe 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*)
X 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*)

<u>Performance Measure Score</u>

Using a mixed effect logistic regression model with varying intercept, a signal to noise analysis was conducted to examine the variability in performance measure score at the plan level, which is the unit of analysis specified

for this measure. This test was conducted with the Medicare PDE data described in sections 1.2- 1.7. This test examines the variance in performance measure score between plans compared to the variance in performance measure score for individuals within plans, and models the individual's SUPD rate based on the varying plan mean. A likelihood-ratio (LR) test was also performed to determine if a model with random effects would fit the data better than a standard logistic regression model without random effects.

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*)

The results of the mixed effect model are outlined in Table 1.

Table 1. Mixed Effect Logistic Regression Model, Statin Use in Persons with Diabetes measure rate comparison across Part D plans

	Coefficient	**Standard**	**Z**	**p-value**
Intercept	1.018	0.009	110.56	<0.001
	Estimate	**Standard**	**95% Confidence Interval**	
Random Effects	0.224	0.008	0.210	0.239

The p-value for the likelihood ratio test was <0.001.

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?*)

Performance Measure Score

The mixed effect logistic regression model results show that the standard deviation of the intercept term is significantly different from zero (0.224), and is also supported by the 95% confidence interval, which does not contain 0. In addition, the likelihood-ratio test shows that the varying intercept model (which allows SUPD rates to vary across contracts) fits the observed data better than a standard logistic regression model without random effects (which restricts all contracts to have the same average SUPD rate) with the p-value of <0.001; significant at alpha=0.05.

These results indicate that there are significant differences in performance measure scores between plans, which allows for discrimination between high performing plans and low performing plans. Based on these results, the SUPD measure is considered to be reliable.

2b2. VALIDITY TESTING

2b2.1. 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**

X **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*)

2b2.2. 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)*

PQA uses a transparent, consensus-based measure development and testing process. The process used in 2014 to develop the measure, Statin Use in Persons with Diabetes (SUPD), is outlined below:

Step 1: The PQA Adherence workgroup identified this measure concept as appropriate for development into a fully specified performance measure during their February 2014 meeting. Workgroups typically focus on specific aspects of the medication-use system and/or specific therapeutic areas for the basis of a measure concept. In this case the workgroup focused on the ACC/AHA Guidelines that recommend diabetic patients age 40-75y/o receive statin medications to reduce their risk of cardiovascular disease. The workgroup included representatives of PQA members with interest in medication use measures and clinical expertise. The workgroup used a consensus-based approach to specify all aspects of the measure concept.

During the process of development, the workgroup invited a guest speaker to provide information about the ACC/AHA guidelines and to answer questions about specific technical specifications in the measure concept as it was being developed. The guest speaker was Joseph Saseen, PharmD Professor in the Departments of Clinical Pharmacy and Family Medicine at the University of Colorado Anchutz Medical Campus. Dr. Saseen is on the Board of Directors for the National Lipid Association and the Board of Pharmacy Specialties, and also is Director of the American College of Clinical Pharmacy Academy Career Advancement program.

After several months of meetings, the SUPD measure concept was developed and recommended (by vote) to the PQA Quality metrics Expert Panel (QMEP) for further evaluation. The voting results (June 23, 2014) by the workgroup were:

34 people voted in favor of recommending the concept to the QMEP
2 people voted against the recommendation
2 people abstained from voting

Step 2: The QMEP reviewed the measure concept on July 2, 2014 to provide an initial assessment of the key properties of performance measures (i.e., feasibility, usability and scientific validity). Measure concepts that are rated highly on these key properties will be tested.

QMEP voted unanimously in favor of testing the measure concept (13-0)

Step 3: The draft SUPD measure was provided to PQA member organizations for their comments prior to preparing technical specifications for pilot testing. The QMEP reviewed the member comments and then reviewed the testing plan based on this all-member feedback. Specific questions addressed in the testing plan included what criteria best identified diabetic patients (denominator) using only prescription claims data and understanding how many people would be in the denominator with contraindications to statin medications.

Step 4: PQA asked two member organizations to test the draft measure. The testing partner implemented the draft technical specifications with their existing datasets and provided a report to PQA that detailed testing results and recommendations for modifications of the technical specifications.

Step 5: The QMEP reviewed the testing results (October 8, 2014) and assessed the feasibility and scientific validity of the draft performance measure, SUPD.

The QMEP vote unanimous in favor of recommending the measure to the PQA membership for endorsement consideration (vote 15-0)

Step 6: Following the QMEP recommendation, the SUPD measure was posted on the PQA web site for member review. Written comments were requested via email, and a conference call for member organizations was held November 5, 2014 to address any questions. This process allows members to discuss their views on the measures in advance of the voting period.

Step 7: PQA member organizations voted on the endorsement of this performance measure in November, 2014. Of the PQA membership that voted, 89% voted to endorse the measure. (56 out of 63 voting member organizations)

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Of the 38 members of the PQA Workgroup that identified and developed the SUPD measure, 89.5% agreed that the measure had face validity, and recommended that the PQA Quality Measures Expert Panel (QMEP) consider it. After reviewing the measure concept specifications and testing results, 100% of the QMEP members recommended the measure to move forward to be endorsed by PQA membership. After review, 89% of PQA members agreed that the measure should be endorsed.

2b2.4. 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?*)

Based upon the rigorous PQA measure development process designed to assure face validity, and the high rate of consensus from the PQA Workgroup, QMEP, and PQA members, this measure has been determined to have face validity.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions—*skip to section 2b4*

2b3.1. 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*)

Patients in hospice are excluded from this measure. No testing was performed on this exclusion, as the data source, prescription claims data, do not contain claims for persons that are in hospice care.

2b3.2. 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*)
N/A

2b3.3. 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*)
N/A

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1./S13What 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

2b4.1.2. If an outcome or resource use 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.

2b4.2/S14. Identify the statistical risk model variables (Name the statistical method – e.g., logistic regression and list all the risk factor variables.

2b4.2.1/S15. Detailed risk model specifications including coefficients, equations, codes with descriptors, definitions(may be attached in an Excel or cvs file)

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics(*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)
N/A

2b4.11.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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To assess significant differences in measure rates, we used 2013 Medicare Part D data for 736 plan contracts, and calculated the distribution mean, median, standard deviation, and interdecile range. These statistics are reported below in 2b5.2, Tables 1 and 2.

Data from the 2012 Truven Health MarketScan® Databases were also analyzed to assess differences between Medicare, Commercial and Medicare plans. Using a test of proportions, a p-value was calculated to determine if there are statistically significant differences in measure rates between health plans. Comparisons were made for differences in rates between Medicare and Medicaid, Medicare and Commercial, and Commercial and Medicaid (2b5.2, Table 3).

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 1. Variation in SUPD Measure Rates -2013 Medicare Part D data

Mean	Median	Standard Deviation
72.45	72.57%	8.31%

Table 2. Interdecile Range of SUPD Measure Rates -2013 Medicare Part D data

10th Percentile	66.07%
20th Percentile	68.49%

30th Percentile	70.45%
40th Percentile	71.53%
50th Percentile	72.57%
60th Percentile	73.46%
70th Percentile	74.81%
80th Percentile	76.61%
90th Percentile	80.23%
100th Percentile	100.00%
Interdecile Range	14.17%

Table 3. SUPD Measure rates, by insurance type – 2012 Truven Health MarketScan® Databases

Insurance Type	Measure Rate	p-value Medicare vs. Medicaid	p-value Medicare vs. Commercial	p-value: Medicaid vs. Commercial
Medicare	73.6%	<0.0001	<0.0001	
Medicaid	59.1%	<0.0001		<0.0001
Commercial	60.4%		<0.0001	<0.0001

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The measure rates for the 2013 data showed significant variation, with a standard deviation of 8.31%. The 10th percentile rate was 66.07%, the maximum rate was 100%, and the interdecile range was 14.17%. The 2012 data demonstrated statistically significant differences in measure rates between heath plan types (P<0.0001 for all comparisons). The measure rates ranged from 59.1% in the Medicaid population to 73.6% in the Medicare population. This variation shows meaningful differences in whether patients aged 40-75 with diabetes are receiving statins.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

- Only one set of specifications is provided for this measure.

Note: *This criterion is directed 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 SDS 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.

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications?(*e.g., correlation, rank order*)
N/A

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications?(i.*e., what do the results mean and what are the norms for the test conducted*)
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

With the utilization of prescription claims data as the data source for this measure, the dispensing information (including medication, days supply and quantity dispensed) is available for each patient. The requirements to utilize this measure include medication (National Drug Code, or NDC), days supply and prescription fill date.

Since each of these data elements are available via prescription claims data, it is not expected—nor was it found—that missing data would result. Age is derived from the date of birth in the enrollment data. The date of birth in the CMS Medicare Enrollment Database (EDB) is considered to largely be valid and reliable since it determines eligibility for enrollment and payment of services.

2b7.2. 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*)

The frequency of missing data is zero for the current analysis.

2b7.3. 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*)

As stated above, no missing data was found through testing, nor would missing data be expected to occur in the future. Therefore, performance results would not be biased, as prescription claims data provides the data elements necessary to calculate the measure rate.

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.**

Other

If other: Prescription claims data

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)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Prescription claims data is required for payment to health plans, so there is no extra burden or cost in the collection of the data. There have been no feasibility issues with the use of this measure.

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)*.

PQA develops and maintains numerous performance measures related to the medication use system. The Measures are the proprietary property of PQA, and it is in the interest of PQA to protect and promote the appropriate use of the Measures. PQA may approve an organization's use of the Measures; however, no organization may use the Measures without first obtaining permission from PQA prior to using the Measures. Certain uses of the Measures are only approved with a licensing agreement from PQA that specifies the terms of use and the licensing fee. PQA reserves the right to determine the conditions under which it will approve and/or license the Measures.

Licenses are granted on a year-to-year basis. PQA reserves the right to audit the licensee's use of the Measures and may revoke a license if it is determined that the licensee has used the Measures in a manner that is outside the scope of permitted use that was specified in the licensing agreement.

Licensees using PQA measures for commercial purposes are required to pay a fee. The licensing fee may be structured as a fixed annual amount or as a variable amount that is dependent on the volume of utilization of the Measures. As a benefit of membership, PQA members who use the Measures only for internal quality improvement initiatives (i.e., self-assessment) will not be assessed a licensing fee.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Regulatory and Accreditation Programs	Quality Improvement (Internal to the specific organization) Medicare Part D Patient Safety Reportsty http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/index.html?redirect=/PrescriptionDrugCovGenIn/06_PerformanceData.asp

4a.1. For each CURRENT use, checked above, provide:

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

Name of program: Medicare Part D Patient Safety Reports
Purpose: Quality improvement and monitoring
Geographic area – National; nearly 700 health plan sponsor contracts

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

The measure is new (endorsed by PQA membership in November 2014) and is being further evaluated by Medicare Part D. It is currently being reported by CMS to all Medicare Part D health plan sponsors in the monthly Patient Safety Reports. The reports are based on 2015 prescription drug event (PDE) information.
URAC is planning to add this measure as an exploratory measure to their accreditation programs for Community Pharmacy, PBM, and Mail Service.

4a.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.)
CMS is considering this measure as a new 2017 display measure (using 2015 data) and as a possible 2018 Star Rating measure (using 2016 data).
Source: Memorandum April 2015: Announcement of Calendar Year 2016 Medicare Advantage Capitation Rates and Medicare Advantage and Part D Payment Policies and Final Call Letter. Attachment VII: 2016 Call Letter Pg 111
http://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/index.html

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. **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.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No negative consequences were incurred by individuals or populations during the testing nor has any evidence of unintended negative consequences to individuals or populations been demonstrated.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

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

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
The following NQF endorsed measure was not included on the drop down above:

NQF Measure 0729: Optimal Diabetes Care- Cholesterol Statin Use Component
Measure Steward: MN Community Measurement

5a. Harmonization

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 completely harmonized?

No

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

Differences between measures 0729 and 2712: The composite measure, 0729, addresses A1c, blood pressure, statin use, tobacco non-use and daily aspirin or anti-platelet use for patients with diagnosis of ischemic vascular disease. Measure 2712 addresses one specific aspect of appropriate medication use, statin medications in a population with diabetes age 40-75. The composite measure, 0729, is reported at the clinician level and uses data from the medical record. Measure 2712 is reported at the health plan level is based on prescription claims data. The composite measure 0729 includes diabetic patients 18-75 years, while measure 2712 only includes diabetic patients age 40-75 years. While the intent and basis of the measures are similar, there are some differences in the measure specification. These differences are due to the accessibility of clinical data for measure 0729 including LDL, allergies, diagnosis etc. Rationale: The rationales of the measures are similar as they address the same guideline but in different settings of care. Impact on interpretability: These measures will be interpreted differently since one (0729) is a composite measure of diabetes care used by clinicians in an ambulatory setting. The other measure (2712) is specific to statin use in a limited age group of diabetics and will be used by health plans and pharmacists. Data collection burden: There will be no additional level of burden as the data used in measure 2712 is prescription claims data and administrative data that are already collected by the health plan.

5b. Competing Measures

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

OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Pharmacy Quality Alliance (PQA, Inc.)
Co.2 **Point of Contact:** Julie, Kuhle, jkuhle@PQAalliance.org, 515-554-6685-
Co.3 **Measure Developer if different from Measure Steward:** Pharmacy Quality Alliance (PQA, Inc.)
Co.4 **Point of Contact:** Julie, Kuhle, jkuhle@PQAalliance.org, 515-554-6685-

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.

Role of the PQA Adherence Workgroup:
PQA is a consensus-based membership organization. PQA Workgroup members represent a diverse group of stakeholders with expertise in clinical, quality improvement and prescription drug data. This measure concept was developed by the PQA Adherence Workgroup in 2014.

The members of the PQA 2014 Adherence Workgroup include:
Ritchie Madeline Academy of Managed Care Pharmacy (AMCP)
Bain Amanda Academy of Managed Care Pharmacy (AMCP)
Biernacki Anne Marie ActualMeds Corporation
CardenThomas Aetna
Markevich Andy Ahold USA
Patel Vaishali Allergan
Mistry Trusha American Association of Colleges of Pharmacy (AACP)
Haydon-Greatting Starlin American Pharmacists Association (APhA)
Capehart Krista American Pharmacists Association (APhA)
Gunter J. Ashley American Society of Health-System Pharmacists (ASHP)
Delaney Evan Amerigroup
Kounelis Peter John AmerisourceBergen Corporation
Davis Carol Anthurium Solutions, Inc. / ASI Services LLC
Kaur Ayesa Applied Research Works
Stacy Jane Astellas Scientific and Medical Affairs, Inc.
Legg Randy AstraZeneca LP
Ayshford Robb Ateb
Conner Suzy Blue Cross and Blue Shield of North Carolina
Witkowski Nancy Boehringer-Ingelheim Pharmaceuticals
Dezii Christopher Bristol-Myers Squibb
Palmieri James California Northstate University College of Pharmacy
Omotayo Yemi Capital Health Plan
Ey Mark CARE Pharmacies Cooperative
Wolf Carolyne Catamaran
Nguyen Michael CenseoHealth
Lee-Martin Alice Centers for Medicare and Medicaid Services (CMS)
Lambert Jennifer Cigna-HealthSpring
Lizotte Margaret CVS Health
Arnold Stephanie Daiichi Sankyo
Serwetman Lea Dovetail Health
Bauman Tina Express Scripts, Inc.
Nowak Jeri Fairview Medication Therapy Management
Matuszewski Karl First Databank
McClelland Scott Florida Blue
Toumadj Ali Gilead Sciences
Miner Paul Gilead Sciences
Lovelace Belinda GlaxoSmithKline
Civin Lynne Gorman Health Group
Lennartz Crystal Health Mart Systems Inc.
Fortuna Laura HealthPartners
Butteri Nicole Highmark Health Services
Young Peinie Humana
Pearce Heather Humana
Frankfort Jim IMS Health
Clelland Carmen Indian Health Services
Kfuri Antoine Inovalon, Inc.
Ziernicki Danielle Johnson & Johnson
Makarem Abir Kaiser Permanente
Hayes Kristin LDM Group
Blank Dawn Lilly USA
Lichucki Rebecca MarkeTouch Media

Eyerly	Sandy	MeadWestvaco
Whalley-Buono	Elizabeth	MeadWestvaco
Logan	Tripp	MedHere Today
Lukoskie	Lynn	Medication Management Systems
Leslie	Scott	MedImpact Healthcare Systems, Inc.
Hogue	Susan	MedVantx, Inc.
Gerhart	Julie	Merck & Co.
Nwachukwu	Ugo	Mirixa Corporation
Rowell	Crescent	National Alliance of State Pharmacy Associations (NASPA)
Masten	Dale	National Association of Chain Drug Stores (NACDS)
Sapp	Aaron	National Association of Chain Drug Stores (NACDS)
Dunklau	Hank	National Community Pharmacists Association (NCPA)
Jester	Laura	National Community Pharmacists Association (NCPA)
Persinger	Gary	National Pharmaceutical Council
Westrich	Kimberly	National Pharmaceutical Council
Wisniewski	Tami	Novo Nordisk, Inc.
Windsheimer	Andrea	Novo Nordisk, Inc.
Burich	Molly	Otsuka America Pharmaceutical, Inc.
Hoopes	Alex	OutcomesMTM
Barrett	Barbara	Parata Systems
Friedman	Steven	PDX, Inc.
Patel	Binal	PerformRx
Gouveia-Pisano	Julie Ann	Pfizer, Inc.
Searle	David	Pfizer, Inc.
Lang	Kelsey	Pharmaceutical Research & Manufacturers of America (PhRMA)
Kelly	Jack	Pharmacist Partners
Scott	Amy	Pharmacy Quality Solutions
Conklin	Mark	Pharmacy Quality Solutions
Erxleben	Tori	PharmMD
Lee	Charles	Polyglot Systems, Inc.
Dauer	Stephanie	Prime Therapeutics
Scanlon	Katie	Publix Super Markets, Inc
Sistrunk	Robin	Publix Super Markets, Inc
Erensen	Jennifer	Purdue Pharma, L.P.
Kahlon	Summer	RelayHealth
McCullough	Jesse	Rite Aid
Bhosle	Monali	RxAnte
SenGupta	Ran	RxPREDICT
McCabe	James	Safeway, Inc.
Romo-LeTourneau	Victoria	Sanofi
Dao	Anthony	SCAN Health Plan
Werner	Shepin	SinfoníaRx
Kebodeaux	Clark	St. Louis College of Pharmacy
Wittbrodt	Eric	Takeda Pharmaceuticals America, Inc.
Losinski	Victoria	Target
Feltman	Matthew	The Kroger Co.
Kirby	James	The Kroger Co.
Lindholz	Colleen	The Kroger Co.
Chabot	Sandye	Therapeutic Research Center (home of Pharmacist's Letter and Prescriber's Letter)
Shipp	Roy	Tri State Distribution, Inc
Miceli	David	Tri State Distribution, Inc
Schilling	Craig	UnitedHealth Group
Hall	Anna	University of Florida College of Pharmacy
Holmes	Erin	University of Mississippi Center for Pharmaceutical Marketing & Management
Keast	Shellie	University of Oklahoma College of Pharmacy - Pharmacy Management Consultants
Daw	Jessica	UPMC Health Plan
Anderson	Janice	URAC

Vargulick	Adam	VoicePort
Garofalo	Tim	voiceTech Inc.
Chazaud	Sandrine	voiceTech Inc.
Medvedeff	David	VUCA Health
Rudkin	Kristi	Walgreen Co.
Marakas	John	Walmart

PQA QMEP members´ role: The PQA Quality Metrics Expert Panel (QMEP) is charged with evaluating the measure concepts proposed by the PQA workgroups and prioritizing the measure concepts for specification and testing. The Panel reviews comments from PQA members on draft measures to determine whether modifications should be made or what variations should be considered during testing. The QMEP reviews the results of the pilot-testing of the draft measures and makes final recommendations to the PQA membership regarding endorsement of the draft measures. The Panel is comprised of persons who have clinical or other technical expertise related to quality measurement. The members are invited to serve on the QMEP by PQA's senior measurement development team. The composition of the QMEP reflects PQA's membership.

Members of the 2014 QMEP include:

Steven Burch	GSK
Catherine Coast	Highmark
Lynn Deguzman	Kaiser Perm
Chris Dezii	BMS
Chris DuPaul	CVS/Caremark
Karen Farris	U of Michigan/APhA
Pat Gleason	Prime Therapeutics
Mary Ann Kliethermes	Midwestern University/APhA
Terri Moore	formerly URAC – historic consultant in 2014
David Nau	PQS
Bimal Patel	MedImpact
Chris Powers	CMS
Kent Summers	Astellas
Mitzi Wasik	Coventry
Jenny Weber	Humana
Keith Widmer	Express Scripts
Gary Young	Northeastern University

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2014

Ad.3 **Month and Year of most recent revision:** 11, 2014

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

Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.6 **Copyright statement:** Rights retained by PQA, Inc. 2015

Ad.7 **Disclaimers:** N/A

Ad.8 **Additional Information/Comments:** N/A

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 #: 2740
Measure Title: Proportion of Patients with coronary artery disease (CAD) that have a Potentially Avoidable Complication (during the episode time window)
Measure Steward: Health Care Incentives Improvement Institute Inc. (HCI3)
Brief Description of Measure: Percent of adult population aged 18 + years who triggered an episode of coronary artery disease (CAD), are followed for at least one-year, and have one or more potentially avoidable complications (PACs). PACs may occur any time during the episode time window. Please reference attached document labeled NQF_CAD_all_codes_risk_adjustment_06.30.15.xls, in the tabs labeled PACs I-9 and PAC I-10 for a list of code definitions of PACs relevant to CAD.
We define PACs as one of two types:
(1) Type 1 PACs - PACs directly related to the index condition: Patients are considered to have a PAC, if they receive services during the episode time window for any of the complications directly related to CAD, such as for hypotension, cardiac arrest, fluid and electrolyte disturbances etc.
(2) Type 2 PACs - PACs suggesting Patient Safety Failures: Patients are also considered to have a PAC, if they receive services during the episode time window for any of the complications related to patient safety failures such as for sepsis, infections, phlebitis, deep vein thrombosis, pressure sores etc..
All relevant admissions in a patient with CAD are considered potentially avoidable and flagged as PACs.
PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook labeled NQF_CAD_all_codes_risk_adjustment_06.30.15.xls serves as an example. The tab labeled PAC overview gives the percent of CAD episodes that have a PAC and the tab labeled "PAC drill down" gives the types of PACs and their frequencies in CAD episodes within this dataset.
The information is based on a two-year claims database from a large regional commercial insurer. The database had over 3.2 million covered lives and over $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.
Developer Rationale:
1d.3.
Each individual PAC, when measured in isolation, provides a very limited picture of the performance of the provider(s) who are managing or co-managing the care of the patient. However, looking at all the PACs that may occur individually or concurrently in a patient with a given episode provides a comprehensive picture of the care received by the patient for that particular condition or illness.

Additionally, the frequency of occurrence of individual PACs may be so low that it may require very high sample sizes from individual providers to achieve any meaningful and reliable comparisons. But aggregating all the PACs into a single quality metric creates meaningful scores that can be compared across providers even with relatively smaller sample sizes.

Additionally, a comprehensive measure is easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

1b.1. Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for

various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 60% of its plan members with CAD incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow- up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.

4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

6) BCBSNC: Blue Cross Blue Shield of North Carolina: https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5. http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/

11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.

14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

Numerator Statement: Outcome: Number of patients who triggered an episode of coronary artery disease (CAD), are followed for at least one-year, and had one or more potentially avoidable complications (PACs) during the episode time window.
Denominator Statement: Adult patients aged 18 years and above who triggered an episode of coronary artery disease (CAD) and are followed for at least one-year.
Denominator Exclusions: Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the CAD measure if they are considered not relevant to CAD care.

Measure Type: Composite, Outcome
Data Source: Administrative claims
Level of Analysis: Clinician : Group/Practice, Clinician : Team
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

1d.1. **Composite Measure Construction:** any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)
Component Measures (if endorsed or submitted for endorsement): n/a. The individual complications are considered measurable components

Is this a MAINTENANCE measure submission? ☐ **Yes** ☒ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: n/a **Most Recent Endorsement Date:** n/a

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for this *health outcomes* measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating clinical evidence asks if the there is a relationship between the measured health outcome and at least one clinical action is identified and if it is supported by the stated rationale. For a composite measure, the developer must discuss the reasoning for the composite quality constructs, the rationale for constructing, & aggregation and weighting of measure components.

- This new risk-adjusted (by age, gender and clinical co-morbidities) outcomes composite measure assesses the proportion (rate) of adult patients with Coronary Artery Disease (CAD) with *at least one* Potentially Avoidable Complications (PAC) within 12 months of CAD triggered claims data. Based on NQF's criteria, this measure is considered an "any or none" composite measure that assesses if 1 or more PACs or "care defects" have occurred for the index episode. For this composite measure, the individual complications considered the measurable components. PACs are classified in two types: 1) related to CAD, and 2) related to Patient Safety Failures. The2 PAC types are combined into a single "any or none" (bimodal "yes" or "no") PAC rate. PACs are considered unwarranted health outcomes that combine concepts from AHRQ PSIs, PQIs and the CMS HACs and episode-specific PACs into index episode all-cause patient harms rate.
- The developer links primary & secondary prevention care gaps, poor patient education, poor care coordination and poor follow-up increase unnecessary ER visits, hospitalizations, readmissions, and mortalities to increased PACs, and state that PACs for CAD patients should occur rarely in well-managed patients.
- The evidence for Patient Safety Failure PACs is described to be within the influence of the measured entity, though the rationale for selecting some of the identified PACs is not clear (e.g., post procedural fever, oral bisphosphonates, hallucinations). The developer provides an extensive list of comorbidities as risk factor for increased PAC potential, though the severity is not captured in consistently within the claims data.
- In addition to linking processes of care to outcomes, the developer provides an extensive PAC literature review in sections 1a.2. and 1a.2.1. for CAD, Patient Safety Failures & processes of care.
- The developer discusses the rationale for constructing, aggregation and equal weighting for the measure.

Questions for the Committee:

- *Does sufficient evidence exist connecting Patient Safety Failures to the CAD index episode?*
- *Does sufficient evidence exist between the measured health outcome and at least one clinical action identified and supported by the stated rationale?*

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

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

- The developer identifies significant CAD prevalence data as the leading cause of US mortalities, as well as BOOST and the Dartmouth Atlas Project CAD readmission rates.
- PAC performance gaps are calculated from PROMETHEUS administrative claims data from April 1, 2012 - December 17, 2014, for providers with ≥ 10 attributable index episodes. The data includes 468 of 5840 (8.0%) providers from 31,093 of 63,972 (48.6%) index episodes in 3,258,706 unique beneficiaries).

Unadjusted PAC Rates:		**Risk-Standardized PAC Rates (RSPR):**	
Median (IQR):	40.0% (29.9%, 54.8%)	Median (IQR):	40.1% (32.6%, 47.5%)
Range:	0% -100%	Range:	0% - 84%

- Descriptive data on the patient, provider and payer are not provided. The developer provides "Overview" and "Drill Down" PAC rates in the spreadsheet demonstrating gap, though PAC rates for individual complications are not provided.
- The developer does not provide data on disparities.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - This is an outcome measure. The developers DO NOT Present any evidence of this measure being performed in pilot or research setting.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- Not Applicable

1b. ***Committee's Comments on Performance Gap:***

- Performance Gap presented for outcomes in CAD but not with this composite measure.

1c. ***Committee's Comments on Composite Performance Measure:***

- I did not find this measure clearly presented with respect to performance.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure assesses the rate of patients with 1 or more PAC(s) during index episodes. This new risk adjusted outcomes measure is specified for use at the clinician group and team levels of analyses.
- The measure exclusively uses electronically available administrative claims data to calculation the measure score, and for this measure, better care equals lower scores.
- The developer states that non-patient-related PACs as controllable by provider processes without further analysis as other influencers that may contribute to PAC rates beyond the patient and provider (e.g., payer, access, suppliers, etc.).
- Patient- and claims- based exclusions are provided to promote the availability and consistency of claims data capture, including payer enrollment requirements, cost outliers of < 1% or > 99%, and claims not relevant to CAD.
- Developers provide administrative claims data for CAD & PAC (CAD- & Patient Safety Failure-related) triggers, and describe a complete 12-month episode time window. A calculation algorithm is provided.
- ICD-9 & ICD-10 codes are provided. ICD-10 descriptions & the ICD-10 conversion methodology are not provided.
- A conceptual risk model and statistical method using logistic regression model for determining the probability of a patient incurring a PAC are provided. After adjusting for patient-related factors, the developers state the remaining

PAC variance is due to factors potentially controlled by the provider during and after hospitalization. "Predicted" coefficients from the risk adjustment models are summed to give predicted probabilities of PAC occurrence.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- The developer tested reliability at the performance measure score, and used a beta-binomial model and a signal-to-noise analysis, which is appropriate for this type of measure, to differentiate the true difference between measured entities (the signal) to random measurement error (the noise). A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in between *provider* performance. A value of 0.7 is often regarded as a minimum acceptable reliability value.
- The measure is specified for CAD patients ≥ 18 years, though the testing sample includes patients 18 through 64 years.
- Providers with < 10 CAD episodes were excluded from reliability testing, though the measure is specified for patient without episode restrictions. A sample of 468 of 5840 (8.0%) providers and 31,093 of 63,972 (48.6%) CAD episodes were included in the testing, for patients with a mean age of 56.4 (18-64 years) with 27% being female.
- The developer states, "Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset will apply to another." The developer also states that very high sample sizes are to achieve any meaningful and reliable comparisons.
- A patient may have more than one condition-specific concurrent episode with claims applied to both episodes. If an inpatient claim corresponds to a procedure index episode and to a condition index episode, the claim would be assigned to the procedure index episode, rather than the condition index episode (e.g., for a claim that corresponds to both index episodes of CAD & CABG, the claim would be assigned to CABG).
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- Reliability results are provided in the table below, as well as in great detail in the accompanied spreadsheet with median (IQR) results demonstrating median reliability of 0.73 (0.61,0.83) for ≥ 10 providers, increasing with the number of episodes per provider. For reliability analysis, providers were restricted to the minimum of 10 CAD episodes, though all episodes were included in the risk model.
- The developer provides a supplementary fact sheet (available for review on SharePoint) requiring a minimum of 21 index episodes for absolute reliability, and a minimum of 10 index episodes for median reliability of > 0.7.

Reliability Scores	Minimum # Episodes Per Provider		
	>=10	>=25	>=50
# of Providers (%)	468 (100)	171 (37)	80 (17)
Median (IQR)	0.73 (0.61,0.83)	0.85 (0.79,0.91)	0.92 (0.88,0.95)
Range	0.50-1.00	0.72-0.99	0.84-0.99

The table provides a summary reliability scores minimum sample size thresholds. Complete results are provided in the workbook entitled, NQF_CAD_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab.

Questions for the Committee:

- *Reliability testing was conducted only for those providers with at least 10 episodes. Can differences in performance*

be identified for providers with fewer than 10 episodes? For patients ≥ 65 years?

- *Should the measure be specified to include only those providers with at least 10 episodes?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- Because this is an outcome measure, the rationale that is presented for subcriterion 1a does not necessarily have to address all of the variables used to calculate the measure.
- The measure uses a statistical risk model with 170 risk factors and episode-specific subtypes/severity markers including age, gender and clinical comorbidities, on at least 10 claims to determine "stable" covariates and assess comorbidity or procedure impact on the PAC. All covariates must be present prior to an episode trigger. No formal covariate analysis was conducted to select risk factors beyond the present on a minimum of 10 claims threshold. The developer describes the heterogeneity of the provided data sets as crucial to ensure measure validity, and the accuracy and completeness of the data sets.
- The developers did not provide disparities data, an exploration of a conceptual relation to SDS, or empirical testing of SDS factors in the risk model.

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Are these variables available and generally accessible for the measured patient population?*

2b2. Validity testing

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.

- The developer conducted systematic assessment of face validity for the performance measure score for validity testing in numerous ways, including the use of monthly multi-specialty clinical working groups and focus groups, and face validity comparisons of the measure to other national accountability measures, as well as additional literature for the measure & PAC development process.
- No empiric results are provided for the face validity tests described above.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer describes patient- (demographic, enrollment or low/high claims cost) and claims-based (due to missing or non-relevant data) exclusions for the measure. They further state nearly half of the original population of CAD patients was removed from the denominator with applied exclusions.
- A significant number of patients were eliminated from the measure due to exclusion criteria, including 31,093 of 63,972 CAD (48.6%) episodes (in 3,258,706 unique beneficiaries) and 468 of 5840 (8.0%) providers.

Questions for the Committee:

- *Are high cost outliers (> 99%) exclusions an opportunity to identify PACs?*
- *Does the high number of exclusions restrict the measure use?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- The risk model (detailed in the accompanied spreadsheet) includes 170 factors and subtypes including age, gender, 12-month enrollment markers, co-morbidities, and episode severity markers.
- No SDS factors beyond age and gender were included in the risk-adjustment approach. The developers note that race was not available for analysis, and no description of the of the conceptual relationships between patient sociodemographic factors, patient clinical factors, quality of care, and the outcomes (PAC rates) were provided, nor do they discuss the availability of SDS variables.
- Logistic regression was used to model the probability of at least one PAC during an episode. The reasoning for no additional modeling performed is described.
- The performance of the model was determined with a split sample method by estimating the model coefficients using a development dataset (80% of the sample) and applying those coefficients to a validation dataset (20% of the sample). C-statistics for the development and validation samples with c-statistic results of 0.803 and 0.792, respectively. C-statistics measures the extent of a statistical model to discriminate between a patient with and without PAC, with an ability to predict if a PAC is or is not present about 75% of the time. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Generally, models with c-statistic values of at least 0.7 are considered good.
- Both Hosmer-Lemeshow Goodness-of-Fit statistics and risk-decide plots were provided to indicate model fit. Results from the Hosmer-Lemeshow test suggest that the fit is not good; however, this test is sensitive to the number of groupings and sample sizes. Results from the risk decile plot indicate that the predicted PAC rates are similar to the observed PAC rates across all deciles of risk. The developer states the model demonstrates strong predictive power.

Questions for the Committee:

- *Is the Committee aware of conceptual relationship(s) between additional patient-level SDS factors, patient clinical factors, quality of care, and PACs (other than gender and age)? If so, what data might be available to allow an empirical analysis of these relationships?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*

2b5. Meaningful difference:

- The developer presents PAC rates across providers and also providers adjusting for differences in patient severity in a ratio of observed to expected attributable episodes to PACS accounting for patient severity, and calculates estimates from the risk model, for risk-standardized PAC rates for the provider.

Summary of Unadjusted and Adjusted Performance Scores Across Providers:

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25
Unadjusted		
Median (IQR)	40% (30%, 55%)	39% (32%, 51%)
Range	0%-100%	8%-94%

Adjusted (RSPR)*		
Median (IQR)	40% (33%, 47%)	41% (34%, 46%)
Range	0%-84%	11%-71%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_CAD_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- As there is only one data source used for measure calculation (administrative claims), comparability of data sources or methods is not applicable.

2b7. Missing Data

- No formal analysis of missing data is provided. As the measure assesses the rate of patients with PACs, rather than the rate of PACs per index episode, the total number of PACs is not included in the PAC rate.
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- The developers state the under-coding of claims is unavoidable in the current DRG payment structure which could lead to under capture or missing PACs.

2d. Empirical Analysis to Support Composite Construction

- As an "any or none" composite, the individual complications are considered measurable components of the composite. Frequency and distribution statistics are provided in the PAC Overview and PAC Drill Down tabs in the measure spreadsheet, which detail PAC types and subtypes. The identification of individual PACs are not provided (e.g., sepsis, unattended falls, DVT).
- PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.
- The developer states that no formal analysis was performed on missing data. For details, see 2b7 above.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- I did not find the Risk methodology clearly defined.

2a2.: ***Committee's Comments on Reliability-Testing:***

- Testing done on a small number of providers

2b1.: ***Committee's Comments on Validity-Specifications***:

- Developer stated face validity but no pilot data presented.

2b2.: ***Committee's Comments on Validity-Testing:***

- No data presented

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Not Applicable

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than the person obtaining the original information (e.g., DRG, ICD-9 codes on claims).
- The developer provides an excel spreadsheet attachment including claims coding for diagnoses, visits, hospitalizations, post-acute facility stays, procedures, laboratory tests and procedures/surgeries, for CAD & PAC triggers, and describe the time window for measuring PACs as 12 months following a CAD episode triggers, as well as a decision tree for measure calculation and implementation.
- The measure specifications, metadata and calculation algorithms are available free of charge on the developer's website. Limited analytics are planned at no cost to the end user.
- This is not an eMeasure.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- I do not think this is feasible in current environment.

Criterion 4: Usability and Use

4. Usability and 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.

- This is a newly developed claims measure is currently used in accountability programs for payers, states, and planned for public reporting, professional certification or recognition programs, and external quality improvement for benchmarking purposes.
- The developer states that PAC measures provide a foundation for the relationship between healthcare quality and cost and assist in the exploration of practice reengineering and alternative payment models, act as indicators of potential harm, and is spurring the development of private-based analytics software for further outcomes exploration. No public improvement rates are available due to recent implementation and variation in PAC definitions have also modified, though the provided PROMETHEUS data suggest wide variation in performance and improvement opportunities. Payer and Provider improvement use perspectives are also outlined.
- The developer stated they identified no unintended consequences, though they also state the measure used in small volumes may be used for QI purposes, though they may produce unreliable performance scores. They further state that under-coding or "gaming" of claims is unavoidable in the current DRG payment structure could be an unintended consequences of the measure, and payers calculating the measures even with inadequate sample sizes and using the results to penalize providers could lead to invalid provider comparisons.
- If the measure was theoretically to be used for accountability purposes to "ding" the measured entity, further exploration of PAC antecedents and the measured entity is warranted, especially with small group practices and very small PAC rates. In the original testing sample of 5840 providers, when providers with fewer than 10 CAD episodes were eliminated from analysis due to less reliable estimates with small numbers, 468 (8.0%) remained for analysis.

Questions for the Committee:

- *Is the measure publicly reported?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

- *Should PAC measures also include the clinician: group in the analysis or include population-level only entities?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- Tools will need to be developed for this to be practical.

Criterion 5: Related and Competing Measures

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

0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2740

Measure Title: Proportion of Patients with coronary artery disease (CAD) that have a Potentially Avoidable Complication (during the episode time window)

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: 6/30/2015

Instructions

- *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.

- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☒ Health outcome: Potentially Avoidable Complications

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

CAD is a chronic condition with a high prevalence rate that can be controlled by primary and secondary prevention, the guidelines for which are well established (Weintraub 2011). Non-compliance to primary and secondary prevention is associated with potentially avoidable complications such as cardiovascular hospitalizations, the need for revascularization procedures and also an increase in all-cause cardiovascular mortality (Mozaffarian 2015). While there is a general understanding of the nature of care failures in chronically ill patients (e.g. ambulatory care sensitive hospitalizations) (Yuen 2004), there has been no attempt to measure the magnitude or the type of potentially avoidable complications, and the cost reductions that would ensue if a payment model encouraged care to be optimized at benchmarks achieved in studies.

There is enough evidence in the literature that highlights significant "gaps in care" in management of patients with chronic conditions (McGlynn 2003). Gaps in care, in turn lead to process failures that cause patients to incur unnecessary services and some harm (Jha 2013). For example, a report by the Agency of Health Care Research and Quality (AHRQ) highlighted the fact that in 2008, $4.4 million out of a total of 39 million (11 percent) hospital-stays that could have been prevented (Stranges 2008); and for Medicare beneficiaries one in five admissions were for a potentially preventable condition (Jiang 2006). To improve accountability in the delivery of medical care, AHRQ has developed a list of patient safety indicators (PSIs) to identify potential harms to patients and a list of ambulatory care sensitive conditions (ACSCs) to identify admissions that could have been potentially avoided with good outpatient care. AHRQ PQI 13 (Patient Quality Indicator) measures admission rates for angina without procedure (AHRQ 2008). Additionally, the Centers for Medicare and Medicaid Services (CMS) have taken a "Six Sigma" approach and defined Hospital Acquired Conditions (HACs) and "never events" that should almost never occur and are applying financial penalties when these events do occur (CMS 2012).

The Potentially avoidable complications (PAC) measure goes beyond the AHRQ PSIs, PQIs and the CMS HACs and creates a single comprehensive measure that measures all-cause harms for a patient with the index condition. Potentially avoidable complications (PACs) are the unwarranted health outcomes that this measure addresses (de Brantes 2010). Lack of patient education on self care techniques, poor care coordination, and poor arrangements of patient follow-up could lead to unnecessary ER visits, hospitalizations and gaps in care leading to increased morbidity and even repeat acute coronary syndrome including acute myocardial infarction (Weaver 2013) (Southern 2014). All these adverse events are aggregated together as a single comprehensive measure to study the overall rate of PACs in the CAD population.

Adult patient diagnosed with Coronary Artery Disease

↓

Physician practices fail to educate patients / Physician practices have poor access

↓

Patient visits ER / gets hospitalized (Ambulatory care sensitive hospitalization event)

↓

Patient discharged with management advise / remains in hospital for treatment of PAC

Well-managed patients with CAD should rarely incur a potentially avoidable complication such as an emergency room visit, and hospitalizations related to CAD should occur only in the rarest of circumstances.

The enclosed workbook entitled NQF_CAD_all_codes_risk_adjustment_06.30.15.xls lists the types of PACs and their frequency as calculated in a large regional database (see tab PAC overview). Over 42% of patients with CAD had a PAC, with about 17% of PACs directly related to CAD itself, such as hypotension, syncope or fluid and electrolyte disorders (see tab PAC Drill Down Graph). Although the preventable hospitalizations in the CAD population were low, at only 3.5% of all CAD episodes; approximately 34% of patients with CAD had PACs related to patient centered care failures such as poor control of diabetes, respiratory insufficiency and acute gastritis, many of them being managed in an outpatient setting in physician offices. As a result over 40% of episodes had a PAC indicator on the professional claims.

While PACs may not be eliminated completely, identifying the magnitude of PACs and knowledge of the cause for the most frequent or the most expensive PACs could place an emphasis in reducing them and as a consequence improving patient outcomes. The ability to clearly identify the type and frequency of each PAC creates a highly actionable measure for all providers that are managing or co-managing the patient; as well as for the health plan with whom the patient is a member (de Brantes 2009).

References:

1) Weintraub, William S., et al. "Value of Primordial and Primary Prevention for Cardiovascular Disease." *American Heart Association* (2011): 1-24. Web.

2) Mozaffarian, Dariush, MD, DrPH, FAHA. et al. "Heart Disease and Stroke Statistics - 2015 Update." *American Heart Association Stastical Update*. American Heart Association, 14 June 2015. Web.

3) Yuen, E. ‘‘Severity of Illness and Ambulatory Care-Sensitive Conditions.’’ *Medical Care Research and Review* 61.3 (2004): 376–91.

4) McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. ‘‘The Quality of Health Care Delivered to Adults in the United States.’’ *New England Journal of Medicine* 348.26 (2003): 2635–45.

5) Jha, A. K., I. Larizgoitia, C. Audera-Lopez, N. Prasopa-Plaizier, H. Waters, and D. W. Bates. "The Global Burden of Unsafe Medical Care: Analytic Modeling of Observational Studies." *BMJ Quality & Safety* 22.10 (2013): 809-15. Web.

6) Stranges, Elizabeth, MS, and Carol Stocks, RN, MHSA. "Potentially Preventable Hospitalizations for Acute and Chronic Conditions, 2008." *Agency for Healthcare Research and Quality* 99 (2010). Web.

7) Jiang HJ, Russo CA, and Barrett ML. Nationwide Frequency and Costs of Potentially Preventable Hospitalization, 2006. AHRQ-HCUP Statistical Brief # 72. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb72.jsp.

8) Department of Health and Human Services, Agency for Healthcare Research and Quality. 2008. "AHRQ Quality Indicators. Prevention Quality Indicators: Technical Specifications, Version 3.2" Available at http://www.qualityindicators.ahrq.gov

9) Centers for Medicare and Medicaid Services. "Hospital Acquired Conditions." *CMS.gov*. Centers for Medicare and Medicaid Services, 2012. Web.

10) de Brantes, Francois, Amita Rastogi, and Michael Painter. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." *Health Services Research 2nd ser.* 45.6 (2010): 1854-871. Web.

11) Weaver, Sallie J., Lisa H. Lubomksi, Renee F. Wilson, Elizabeth R. Pfoh, Kathryn A. Martinez, and Sydney M. Dy. "Promoting a Culture of Safety as a Patient Safety Strategy." *Annals of Internal Medicine* 158.5_Part_2 (2013): 369-75. Web.

12) Southern, D. A., J. Ngo, B. J. Martin, P. D. Galbraith, M. L. Knudtson, W. A. Ghali, M. T. James, and S. B. Wilton. "Characterizing Types of Readmission After Acute Coronary Syndrome Hospitalization: Implications for Quality Reporting." *Journal of the American Heart Association 3*.5 (2014): 1-8. *PubMed*. Web.

13) de Brantes, François M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. "Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model." *NEJM* (2009) 361:1033 (Perspective)

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Rationale: Better processes of care create an atmosphere of proactive management, consistency in care and standardized care patterns (Shekelle 2013) (Fenter 2006). Patient education and adopting safe practices significantly reduces occurrence of potentially avoidable complications (PACs) in all settings (Klein 2011) (Wachter 2013) (Berwick 2006) (Kovner 2011) (Farley 2013). It is known that by holding providers accountable for occurrence and costs of PACs, a built-in warranty is created around care of the index condition (de Brantes 2009).

Specifically, management of CAD using evidence-informed guidelines lead to significantly improved outcomes. A study evaluating the impact of prevention on reducing the burden of cardiovascular disease found that, for patients with established CAD, appropriate LDL cholesterol control could lead to an absolute reduction of MI risk of 40% (Kahn 2008).

Additionally, evidence-based pharmacological therapy such as proper use of angiotensin-converting enzyme inhibitors in coronary artery disease patients leads to reduction of cardiovascular endpoints such as death, MI and strokes (Danchin 2006). Adherence to primary and secondary guidelines for CAD management could lead to better outcomes, and reductions in PACs stemming from CAD (Smith 2011).

Like other chronic illnesses, coronary artery disease is marked by episodes of acute exacerbation requiring hospitalization, most commonly for acute coronary syndromes (ACS). Despite improvements in acute care and survival after ACS hospitalization, early readmissions remain common, and have significant clinical and financial impact (Southern 2014). Preventable readmissions have become a focus of national quality improvement effort.

Studies have demonstrated where care coordination exists, ambulatory care-sensitive hospitalizations decreased by 30% (Bodenheimer 2008). However, if patients do get hospitalized, discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013) (Mittler 2013). Another study from the Boston Medical Center, demonstrated that although one in five hospitalizations are complicated by post-discharge adverse events, development of a strong discharge services program for patients admitted for medical conditions reduced hospital utilization within 30 days of discharge (Jack 2009). In addition, while in the hospital, safe practices reduce the burden of healthcare associated complications (Ranji 2007). Some of these are listed below:

1. Frequent hand-washing reduce hospital acquired infections (WHO 2007)
2. Carefully implemented protocols lead to reduced line sepsis (Pronovost 2010)
3. Aspirin on Discharge prevents repeat AMIs (Hall 2014)
4. Discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013)
5. DVT prophylaxis in patients on bed rest avoids pulmonary embolism (Shekelle 2013)
6. Frequent change in position of CAD patients in the CCU avoids pressure sores (Shekelle 2013)
7. Adherence to primary and secondary prevention guidelines for CAD (Smith 2011)

PAC measures in the setting of CAD look at all-cause harms, such as the ones highlighted above, arising from poor management of a patient with CAD.

References:

1) Shekelle, Paul G., et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

2) Fenter TC, et al. “The cost of treating the 10 most prevalent diseases in men 50 years of age or older.” *Am J Manag Care* 12.4 (2006): S90-S98.

3) Klein S. and D. McCarthy, OSF HealthCare: “Promoting Patient Safety Through Education and Staff Engagement.” *The Commonwealth Fund* 1475.7 (March 2011).

4) Wachter, Robert M., Peter Pronovost, and Paul Shekelle. "Strategies to Improve Patient Safety: The Evidence Base Matures." *Annals of Internal Medicine* 158.5_Part_1 (2013): 350-52. Web.

5) Berwick, D.M., D. R. Calkins, C. J. McCannon et al., "The 100,000 Lives Campaign: Setting a Goal and a Deadline for Improving Health Care Quality," *Journal of the American Medical Association* 295.3 (2006):324–27

6) Kovner, Anthony R., James Knickman, and Steven Jonas. "Chapter 11." *Jonas & Kovner's Health Care Delivery in the United States*. 10th ed. New York: Springer Pub., 2011. Print.

7) Farley, TG, et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

8) de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? *Health Affairs (Millwood)* 28 (2009): w678-w687.

9) Kahn R, Roberston RM, Smith R, Eddy D. "The impact of prevention on reducing the burden of cardiovascular disease." *Circulation* 118.5 (2008):576-585.

10) Danchin N, Cucherat M, Thuillez C, et al. "Review: ACE inhibitors reduce mortality and cardiovascular endpoints in stable coronary artery disease." *Arch Intern Med* 166 (2006): 787-96.

11) Smith, Sidney C., et al. "AHA/ACCF Secondary Prevention and Risk Reduction Therapy for Patients With Coronary and Other Atherosclerotic Vascular Disease: 2011 Update." *Journal of the American College of Cardiology* 58.23 (2011): 2432-446. Web.

12) Southern, D. A., et al. "Characterizing Types of Readmission After Acute Coronary Syndrome Hospitalization: Implications for Quality Reporting." *Journal of the American Heart Association* 3.5 (2014): 1-8. *PubMed*. Web.

13) Bodenheimer, T"Coordinating Care — a Perilous Journey through the Health Care System," *New England Journal of Medicine* 358.10 (2008): 1064-1071.

14) Weaver, Sallie J., Lisa H. Lubomksi, Renee F. Wilson, Elizabeth R. Pfoh, Kathryn A. Martinez, and Sydney M. Dy. "Promoting a Culture of Safety as a Patient Safety Strategy." *Annals of Internal Medicine* 158.5_Part_2 (2013): 369-75. Web.

15) Mittler, Jessica N., Jennifer L. O'hora, Jillian B. Harvey, Matthew J. Press, Kevin G. Volpp, and Dennis P. Scanlon. "Turning Readmission Reduction Policies into Results: Some Lessons from a Multistate Initiative to Reduce Readmissions." *Population Health Management 16*.4 (2013): 255-60. Web.

16) Jack BW, Chetty VK, Anthony D et al. "A Reengineered Hospital Discharge Program to decrease Rehospitalization: A randomized trial." *Ann Int Med* 150 (2009):178-187.

17) Ranji SR, Shetty K, Posley KA, et al. "Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol 6 Prevention of Healthcare-Associated Infections) Technical Review 9" *AHRQ* 04(07)-0051-6. January 2007.

18) The Joint Commission, Aide Joint Commission International, and WHO. "Improved Hand Hygiene to Prevent Health Care-Associated Infections." World Health Organization. WHO Collaborating Centre for Patient Safety Solutions (May 2007). Web.

19) Pronovost, P.J., G. A. Goeschel, E. Colantuoni et al., "Sustaining Reductions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study," BMJ 340:c309 (Feb. 4, 2010): Web.

20) Hall, Hurst M., MD, James A. De Lemos, MD, and Jonathan R. Enriquez, MD. "Contemporary Patterns of Discharge Aspirin Dosing after Acute Myocardial Infarction in the United States: Results from the National Cardiovascular Data Reg." *Circulation: Cardiovascular Quality and Outcomes. American Heart Association Journals*, 12 Aug. 2014. Web.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

Review of literature

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

2740_CAD_Evidence_Attachment_HCI3-635717847016609999.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 60% of its plan members with CAD incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow- up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.

4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.
6) BCBSNC: Blue Cross Blue Shield of North Carolina:
https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5.
http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/

11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.

14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The data included two years of administrative claims covering the period April 1, 2012 through December 17, 2014. There were a total 31,093 episodes of CAD.

Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. After this exclusion 468 (out of 5840) providers remained. Performance scores of these providers are summarized in the following table:

Unadjusted PAC Rates:

Median (IQR):	40.0% (29.9%, 54.8%)
Range:	0% -100%

Risk-Standardized PAC Rates (RSPR):

Median (IQR):	40.1% (32.6%, 47.5%)
Range:	0% - 84%

Please refer to the NQF_CAD_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

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.**

Despite the implementation of evidence-based programs such as the Care Transitions Intervention and Project BOOST at several hospitals, available evidence suggests that readmission rates in patients with CAD remain largely unchanged (Hansen 2011) (Axon 2013). Dartmouth Atlas Project analyzed 2008-2010 data and found that Medicare readmission rates from CAD remained unchanged in more than 90% of academic medical centers (Dartmouth Atlas Project 2013). A study by Brock et al. showed that though community interventions showed a modest reduction in hospital admissions of Medicare patients, readmissions as a percentage of hospital admissions did not change (Brock 2013).

At discharge, patients of CAD should undergo appropriate interventions including secondary prevention as per guidelines issued by the AHA in 2011 to prevent progression of disease leading to increased risk of readmissions (Smith 2011).

The PAC measures go beyond simple readmission rates and look for all-cause harms in patients with coronary artery disease. While PACs may not be eliminated completely, identifying the magnitude of PACs and knowledge of the cause for the most frequent or the most expensive PACs could place an emphasis in reducing them and as a consequence improving patient outcomes. The ability to clearly identify the type and frequency of each PAC creates a highly actionable measure for all providers that are managing or co-managing the patient; as well as for the health plan with whom the patient is a member (de Brantes 2010).

References:

1) Hansen, LO, et al. "Interventions to reduce 30-day rehospitalization: a systematic review". Ann Intern Med 155. (2011): 520-528.

2) Axon, R. N., and E. A. Coleman. "What Will It Take to Move the Needle on Hospital Readmissions?" American Journal of Medical Quality 29.4 (2013): 357-59. Web.

3) Dartmouth Atlas Project, and PerryUndem Research & Communications. "The Revolving Door: A Report on U.S. Hospital Readmissions." The Revolving Door: A Report on U.S. Hospital Readmissions. Robert Wood Johnson Foundation, Feb. 2013. Web.

4) Brock J, Mitchell, et al. "Association between quality improvement for care transitions in communities and rehospitalizations among Medicare beneficiaries." JAMA 309. (2013):381-391.

5) Smith, Sidney C., et al. "AHA/ACCF Secondary Prevention and Risk Reduction Therapy for Patients With Coronary and Other Atherosclerotic Vascular Disease: 2011 Update." Journal of the American College of Cardiology 58.23 (2011): 2432-446. Web.

6) de Brantes, F., A. Rastogi, and M. Painter. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach". Health Services Research 45.6.2 (2010): 1854-1871.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

Not applicable

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

Not applicable

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;

OR

- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.
According to the National Center for Health Statistics 2011, cardiovascular disease represented the leading cause of mortality in the US across genders and ethnic groups. Cardiovascular disease accounts for more deaths than the next four leading diseases in the United States combined. Coronary artery disease (CAD), in particular, accounted for 405,209 deaths in 2008 in the US alone. Moreover in 2010, CAD was projected to cost the US upwards of $108.9 billion dollars (Rimmerman 2015).

CAD is a chronic condition with a high prevalence rate that can be controlled by primary and secondary prevention, the guidelines for which are well established (Weintraub 2011). Non-compliance to primary and secondary prevention is associated with potentially avoidable complications such as cardiovascular hospitalizations, the need for revascularization procedures and also an increase in all-cause cardiovascular mortality (Mozaffarian 2015). There is considerable positive economic impact that can be achieved by reducing the burden of hospital admissions attributable to CAD, by better adherence to care guidelines by both the provider and the patient. Strategies to promote population level health, targeting both individuals at all risk levels for CV disease as well as promoting healthy behavior amongst the people in the community as a whole are important to contain this global epidemic of cardiovascular disease (Mozaffarian 2015).

To improve accountability in the delivery of chronic care, AHRQ has developed a list of prevention quality indicators (PQIs) to identify ambulatory care sensitive conditions (ACSCs) and to measure rates of admissions that could have been potentially avoided with good outpatient care (AHRQ 2008). PQI 13 measures admission rates for angina without procedure. Even though hospitalizations for CAD should be potentially avoidable in their own right; once they do occur, the index stay itself may have a potentially avoidable complication (PAC) or patients may develop a PAC during the post-discharge period. PACs lead to significant variability in outcomes including prolonged hospitalizations, readmissions and emergency room visits, all indicating poor outcomes that harm the patient, cause payers to incur unnecessary costs; and could be improved by providers (deBrantes 2011) (Yong 2010). For coronary artery disease in particular, there is a variety of avoidable complications that could result in a readmission. These complications tend to have a high impact due to the subsequent care needed and associated costs (Smith 2011) (Southern 2014).

Therefore, there are many areas where improvement is possible in CAD, making it a high priority aspect of health care. The PAC measures go beyond simple readmission rates and look for all-cause harms in patients with coronary artery disease.

1c.4. Citations for data demonstrating high priority provided in 1a.3
1) Rimmerman, Curtis M. "Coronary Artery Disease." Cleveland Clinic: Center for Continuing Education. (Feb. 2013). Cleveland Clinic. Web.

2) Weintraub, William S., et al. "Value of Primordial and Primary Prevention for Cardiovascular Disease." American Heart Association (2011): 1-24. Web.

3) Mozaffarian, Dariush, MD, DrPH, FAHA. et al. "Heart Disease and Stroke Statistics - 2015 Update." American Heart Association Stastical Update. American Heart Association, 14 June 2015. Web.

4) Department of Health and Human Services, Agency for Healthcare Research and Quality. 2008. ‘‘AHRQ Quality Indicators. Prevention Quality Indicators: Technical Specifications, Version 3.2’’

5) de Brantes F, Rastogi A, and Sorensen CM. “Episode of Care Analysis Reveals Sources of Variation in Costs.” Am J Manag Care 17.10 (2011): e383-e392.
6) Young, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

7) Smith, Sidney C., et al. "AHA/ACCF Secondary Prevention and Risk Reduction Therapy for Patients With Coronary and Other Atherosclerotic Vascular Disease: 2011 Update." Journal of the American College of Cardiology 58.23 (2011): 2432-446. Web.

8) Southern, D. A., et al. "Characterizing Types of Readmission After Acute Coronary Syndrome Hospitalization: Implications for Quality Reporting." Journal of the American Heart Association 3.5 (2014): 1-8. PubMed. Web.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable

1d. Composite Quality Construct and Rationale

1d.1. **A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
  - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient); or
  - any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient).

1d.1. Please identify the composite measure construction: any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

1d.2. **Describe the quality construct, including:**

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

We classify PACs into two types: Type 1 PACs are directly related to the index condition and are often controlled by the servicing provider; Type 2 PACs, on the other hand result from patient safety failures and could be reduced by better systems and better processes in care. Both types of PACs could occur in any setting and so could be identified through any type of claims coming in the administrative dataset, including in-patient, out-patient, or professional claims. PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1.

The PAC measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs), Hospital Inpatient Quality Reporting measures, Avoidable Readmissions, AHRQ defined patient safety indicators (PSIs), NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates.

All defined PACs, irrespective of their type, or site of occurrence, are aggregated to create an overall comprehensive, composite measure. They all have equal weighting, since they are measured simply by the frequency of their occurrence.

1d.3. **Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.**
Each individual PAC, when measured in isolation, provides a very limited picture of the performance of the provider(s) who are managing or co-managing the care of the patient. However, looking at all the PACs that may occur individually or concurrently in a patient with a given episode provides a comprehensive picture of the care received by the patient for that particular condition or illness.

Additionally, the frequency of occurrence of individual PACs may be so low that it may require very high sample sizes from individual

providers to achieve any meaningful and reliable comparisons. But aggregating all the PACs into a single quality metric creates meaningful scores that can be compared across providers even with relatively smaller sample sizes.

Additionally, a comprehensive measure is easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

1d.4. **Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**
In constructing the comprehensive composite PAC measure, each component PAC, as clinically defined by the subject matter experts, was given the same weight so that arbitrary weights may not bias the results. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence. As such, the patient is the ultimate unit of measurement and if the patient incurred any PAC during the episode, then that counts against the numerator.

Since the emphasis of the PAC measure was to simply identify the occurrence of PACs in any setting, aggregation of the PAC counts to create a comprehensive quality score with equal weights has been met with overall support from the clinical working groups as well as from the implementation sites.

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 subcriteria 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 : Ischemic Heart Disease, Coronary Artery Disease

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Healthcare Associated Infections, Safety : Medication Safety, Safety : Readmissions, Safety : Venous Thromboembolism

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.)*
http://www.hci3.org/ecr_descriptions/ecr_description.php?version=5.2.006&name=CAD&submit=Submit

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_CAD_all_codes_risk_adjustment_06.30.15-635719625915810933.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Outcome: Number of patients who triggered an episode of coronary artery disease (CAD), are followed for at least one-year, and had one or more potentially avoidable complications (PACs) during the episode time window.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The time window is the most recent 12 months of the episode, once a patient has triggered a CAD episode.

S.6. **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, 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.
Patients that have triggered a CAD episode, and are identified as having services for potentially avoidable complications (PACs), during the most recent 12 months of the episode. The enclosed excel workbook entitled NQF_CAD_all_codes_risk_adjustment_06.30.15 gives the detailed codes for PACs in the tabs entitled PACs I-9 and PACs I-10. PACs are identified only based on diagnosis codes.

Services for PACs are identified as follows:
a.Any service (professional, outpatient facility, ancillary) that is relevant to CAD and has a PAC code in any position on the claim
b.Any admission to an acute care facility, that is relevant to CAD
c.Any admission to a post-acute care facility that is relevant to CAD and has a PAC code in any position on the claim

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Adult patients aged 18 years and above who triggered an episode of coronary artery disease (CAD) and are followed for at least one-year.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Please refer to the enclosed excel workbook entitled
NQF_CAD_all_codes_risk_adjustment 06.30.15

The target population is identified using the following criteria:
1.Using administrative claims database, patients with CAD are identified using one of two of the following criteria:
a.Patients having an office visit with a trigger code of CAD in any position, followed by a second confirmatory office visit (with a trigger code of CAD in any position), at least 30 days apart.
b.Patients with a Principal Dx of a CAD trigger code on an in-hospital stay claim.
The trigger codes for CAD are provided in the tab called “Triggers I-9” or “Triggers I-10”.
2. The patient should have continuous enrollment for the entire time window with no more than 30 days as an enrollment gap, with the entity providing the data (so we can ensure that the database has captured most of the claims for the patient during the episode time window).
3. The patient should have a complete episode time window in the claims data – so there are at least 12 months of claims in the database for the patient.
4. Patient should be at least 18 years of age

Once the episode is triggered all relevant claims are assigned to the episode. Relevant claims could be inpatient facility claims, outpatient facility claims, professional services, laboratory services, imaging services, ancillary claims, home health, durable medical equipment as well as pharmacy claims across the entire continuum of care centered around the patient’s episode of care.

Relevant claims are identified as those that have a diagnosis code that matches the codes in the typical Dx codes tabs (Typical Dx I-9 or Typical Dx I-10), or in the PAC Dx codes tab (PACs I-9 or PACs I-10) AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant admissions to acute and post-acute care facilities are also included in the denominator.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the CAD measure if they are considered not relevant to CAD care.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Denominator exclusions include exclusions of "patients" as well as "claims" not relevant to CAD care. Please refer to the enclosed excel workbook entitled (NQF_CAD_all_codes_risk_adjustment 06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
d. If the patient does not have at least 12 months of claims in the database (this helps eliminate incomplete episodes).
e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure if they meet one of the following criteria:
a. If none of the diagnosis codes on the claim are on the list of "triggers" or relevant diagnosis codes (either typical Dx or PAC Dx) for CAD.
b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for CAD.
c. If the CAD trigger hospitalization also triggers a major surgical procedure such as coronary bypass procedure or angioplasty, suggesting that CAD may be a comorbidity or an indication for the surgery.
d. The "principal" diagnosis on an inpatient stay claim during the episode time window triggers its own episode
e. The procedure code on a claim during the episode time window triggers its own episode

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
None

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Conceptual Model
Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, the remaining variance in PACs are due to factors that could be controlled by all providers that are managing or co-managing the patient, both during and after hospitalization.

Statistical Method:
Logistic Regression model to determine the probability of a patient incurring a PAC

Demographic variables, comorbid conditions, as well as clinical severity indicators are fed as independent risk factors into the model. Risk Factors are collected historically. Subtype information is collected from the index claim and any look-back period, if relevant. Subtypes are clinical severity indicators suggesting severity of the episode itself, for example, diagnosis of unstable angina in a CAD patient. For each patient the "predicted" coefficients from the risk adjustment models are summed to give the predicted probabilities of the occurrence of a PAC.

Risk Factors :(Please refer to the enclosed excel workbook entitled (NQF_CAD_all_codes_risk_adjustment 06.30.15.xls). The risk factors along with their codes are listed in the tabs called "All Risk Factors I-9" and "All Risk Factors I-10" and also listed below:

AGE CONTINUOUS VARIABLE
GENDER FEMALE = 1 (MALE IS REFERENCE = 0)

Risk Factor #	Risk Factor Name
RF0101	Anoxic Brain Damage, persistent vegetative state
RF0102	Delirium, Meningitis, Encephalitis
RF0103	Previous Stroke, Paralysis
RF0104	Cerebral Palsy and Other Paralytic Syndromes
RF0105	Spinal Cord Disorders/Injuries
RF0106	Polyneuropathy
RF0107	Multiple Sclerosis
RF0108	Convulsions, Epilepsy
RF0109	Dementia
RF0110	Parkinson´s and Huntington´s Diseases
RF0111	Cerebrovascular Disease
RF0115	after care, rehabilitation
RF0201	visual loss, blindness, retinal tear, detachment
RF0301	ENT, Upper Respiratory Problems
RF0401	Respiratory Failure, O2, ventilator dependence
RF0402	Advanced COPD, Asthma
RF0403	Empyema, bronchiectasis, Pneumonias
RF0404	Aspiration Pneumonia, Laryngeal Problems
RF0406	TB, Pneumoconiosis, Aspergillosis
RF0407	Tobacco use, Lung disease due to External Fumes
RF0408	Other Lung Disease
RF0501	Previous Shock, Syncope, Vent Fibrillation
RF0503	Advanced CHF
RF0504	Cardiomyopathy, valve disorders
RF0505	Cardiac Arrhythmias, Heart Block
RF0506	Pacemaker, AICD
RF0507	Endocarditis, Other post surgical cardiac problems
RF0508	Other Cardiovascular Disease
RF0511	DVT, Pulm Embolism, Pulm Heart Disease
RF0512	Unstable Angina
RF0513	Hypotension, chronic, orthostatic
RF0514	Hyperlipidemia
RF0515	Intraaortic Balloon Pump
RF0516	ventricular assist device, ecmo, prolonged bypass
RF0517	Previous electrophysiology studies, cryoablation
RF0518	Recent AMI
RF0519	Previous PCI
RF0520	Previous CABG
RF0521	Previous Heart & Valve Surgery
RF0522	Previous aortic reconstruction
RF0523	Previos carotid endarterectomy
RF0524	Aortic and peripheral vascular disease
RF0525	Advanced Aortic and Vascular Disease

RF0601 GI Bleed
RF0602 Intestinal Obstruction/Perforation
RF0603 Acute Gastritis, Duodenitis
RF0604 Gastroduodenal Ulcer
RF0606 Intestinal Uro-genital Fistula
RF0607 Abdominal hernia w complications
RF0608 Vascular insufficiency of intestine
RF0609 Inflammatory Bowel Disease
RF0610 Irritable Bowel
RF0611 Diverticulitis, Meckel´s
RF0612 Digestive congenital anomalies
RF0613 Intestinal infection
RF0614 Esophageal Perforation, Hmg, Barretts, Compl Hiatal Hernia
RF0615 Abnormal weight loss
RF0616 Achalasia, Esophageal spasm, Stricture, Dysphagia
RF0617 GERD, Hiatal Hernia, Other Upper GI Disorders
RF0618 Previous Bariatric Surgery
RF0619 Hx of colon polyps, family Hx of colon cancer
RF0620 Enterostomy, GI devices, lap band
RF0701 Pancreatic Disease
RF0702 Perforation, fistula GB, bile duct, pancreas
RF0703 Gall stones, cholecystitis
RF0704 End-Stage Liver Disease
RF0705 Hepatitis, Cirrhosis, Other Hepatbiliary Disorders
RF0706 Recent Gall Bladder, Hepatobilary Surgery
RF0707 Acute Pancreatitis, pseudo cyst
RF0801 Bone/Joint/Muscle Infections/Necrosis
RF0802 Muscular Dystrophy
RF0803 Osteoporosis, ostetits deformans, pathological fracture
RF0804 Rheumatoid Arthritis and Inflammatory Connective Tissue Disease
RF0805 Gout and other crystal arthropathies
RF0806 Other arthropathies
RF0807 Osteoarthritis
RF0808 Joint Deformities
RF0809 Knee derangements
RF0810 Traumatic Dislocation Knee
RF0811 Dislocation Hip
RF0812 Synovitis, Ruture Tendon
RF0813 Status Knee Replacement
RF0814 Status Total Hip Replacement
RF0901 Decubitus Ulcer
RF0902 Skin and wound problems
RF1001 Diabetes, poor control
RF1002 Advanced diabetes
RF1003 diabetes
RF1101 Acute renal failure
RF1102 Dialysis Dependent
RF1103 Nephritis
RF1104 Chronic renal failure
RF1105 Urinary Tract Infections
RF1301 Endometriosis
RF1302 Fibroid uterus, benign tumors of female organs
RF1303 Pelvic Inflammatory disease
RF1304 Uterine prolapse, cystocele, vaginocele
RF1305 Female Harmonal Disorders
RF1306 Ovarian, Broad Ligament Disorders
RF1308 Other disorders of uterus, cervix

RF1309 Menopausal Disorders
RF1310 Menstrual Disorders
RF1401 Multiparity, multigravida
RF1402 Elderly Primi, other
RF1403 Poor obstetric history
RF1406 Cervical incompetence
RF1407 Abnormalities of uterus, female genital tract
RF1408 Hypertension, pre-eclampsia in Pregnancy
RF1409 Severe pre-eclampsia w HTN, Eclampsia
RF1410 Maternal, gestational diabetes, large for date
RF1411 Genital Herpes
RF1412 Infections of genitourinary tract, venereal disease in pregnancy
RF1413 Infectious Diseases in Mother
RF1414 Cardiovascular disease in Mother
RF1415 Mental Disorders in Mother
RF1416 Epilepsy in Mother
RF1417 Liver and biliary tract disorders in mother
RF1418 Kidney Disease in Mother
RF1419 Other Maternal conditions
RF1421 Cephalopelvic Disproportion due to maternal causes
RF1436 Peripartum Cardiomyopathy
RF1441 Previous Cesarean section
RF1450 Maternal Obesity, previous Bariatric Surgery
RF1454 Previous Rupture Uterus, Obstetrical Trauma
RF1458 Complicated Pregnancy Delivery
RF1460 Thrombophlebitis, DVT during Pregnancy
RF1461 Puerperal Sepsis, other major puerperal complications
RF1462 Obstetrical Embolism, Air, Amniotic Fluid, Pulm, Pyemic
RF1467 Tobacco Use in Mother
RF1601 Bleeding Disorders
RF1602 Severe Hematological Disorders
RF1603 Disorders of Immunity
RF1604 Nutritional and other Anemias
RF1605 Long-term use of anticoag, Aspirin
RF1701 Head and Neck Cancers
RF1702 Lung and Intrathoracic Cancers
RF1703 Neuroendocrine, Myeloproliferative Cancers
RF1704 Poorly differentiated, Secondary, Metastatic Cancers
RF1705 Other Tumors
RF1706 Acute Leukemia
RF1707 Cancer uterus, localized female organs
RF1708 Colorectal, Hepatobiliary and other GI cancers
RF1709 Breast, Prostate, Thyroid cancers
RF1710 Testicular Cancer and localized of male organs
RF1711 Cancer of Bladder and Urinary Tract
RF1712 Musculoskeletal Cancers
RF1801 Sepsis, MRSA, Opportunitistic infections
RF1901 Schizophrenia
RF1902 Major Depressive, Bipolar, and Paranoid Disorders
RF2001 Drug/Alcohol Psychosis
RF2002 Drug/Alcohol Dependence
RF2101 Drug Reactions, long term use of drugs
RF2102 Intra-abdominal injury
RF2201 Extensive Third-Degree Burns
RF2301 Major Organ Transplant Status
RF2302 Artificial Openings for Feeding or Elimination
RF2303 Complications of Medical & Surgical Care and Trauma

RF2304 severe morbid obesity
RF2305 morbid obesity
RF2306 obesity
RF2307 mild sleep apnea, hypoventilation
RF2308 moderate sleep apnea, hypoventilation
RF2309 obstructive sleep apnea
RF2310 Severe Protein-Calorie Malnutrition
RF2311 Mild-mod malnutrition
RF2401 Severe Head Injury
RF2402 Major Head Injury
RF2403 Vertebral Fractures without Spinal Cord Injury
RF2404 Falls, Fractures
RF2405 Amputation
RF2501 HIV/AIDS

Subtypes for CAD
Previous CABG, PCI
Unstable angina

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Please refer to the enclosed excel workbook entitled (NQF_CAD_all_codes_risk_adjustment 06.30.15.xls).

Assembling the Denominator:

Using administrative claims database, patients with CAD are identified using one of two criteria: 1) Patients having an office visit with a trigger code of CAD in any position, followed by a second confirmatory office visit (with a trigger code of CAD in any position), at least 30 days apart, 2) Patients a Principal Dx of a CAD trigger code on an in-hospital stay claim. The trigger codes for CAD are provided in the tab called “Triggers I-9” or “Triggers I-10”.

Patients are retained if they are 18 years of age or more, do not have a missing gender, have at least 12 month of claims in the database, have a maximum of 30-day enrollment gap for the entire episode time window, and have no outlier episode costs. All relevant professional, laboratory, imaging, ancillary and other claims that are incurred during the episode time window are included as part of the episode. Claims are considered relevant to CAD care if they have one of the diagnosis codes, as listed on the tab entitled Triggers I-9, Triggers 1-10, PACs I-9, PACs I-10, Typical Dx I-9, or Typical Dx I-10 in any position on the claim AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant admissions to acute and post-acute care facilities are also included in the denominator. All relevant pharmacy claims carrying codes that match the ingredients listed in the Pharmacy tab of the enclosed workbook are also included as part of the episode. Services are pulled as part of the CAD episode based on the diagnosis codes as defined above or if they have a service code that is marked as “sufficient”

for that episode.

If a patient has more than one concurrent episode, and the claim is relevant to both episodes, the claim could get multi-assigned, except in the case of procedural episodes that get carved out with respect to the index stay. So if an inpatient stay claim carried a principal Dx code that matched the trigger diagnosis code for CAD but they also had a procedure code for CABG (coronary artery bypass surgery), the stay claim would get uniquely assigned to CABG and not be counted with CAD.

Once all the episodes are assembled, episodes that match the exclusion criteria, such as those with outlier costs, are flagged (those with total episode costs less than 1st percentile or greater than 99th percentile), and excluded from the final analysis.

Assembling the Numerator:

For every episode included in the denominator, services are flagged as having a PAC (potentially avoidable complication) based on the criteria listed below:

Any service (professional, outpatient facility, ancillary) with a PAC code in any position on the claim

Any admission to an acute care facility, that is relevant to CAD

Relevant claims that do not have any PAC codes, and do not qualify as a PAC based on the criteria outlined above, are listed as typical claims. All included relevant pharmacy services are flagged as typical. Patients that have even a single PAC claim are counted as part of the numerator.

Calculating the measure:

Proportion of CAD patients that have PACs is simply the ratio of patients with PACs within the HTN population and is called the PAC rate as shown in the equation below:

PAC rate = Patients with CAD that have at least one PAC claim / Total number of CAD patients

A flow chart demonstrating the series of steps and the counts of patients at each step is shown in tab entitled Decision Tree of the enclosed workbook called NQF_CAD_all_codes_risk_adjustment 06.30.15.xls

Drill Down Calculations:

Further analysis from this construct helps create actionable reports.

For example as shown in the tab labeled PAC overview, not only do we have the PAC rate for a population, we can break them down by the PAC type – type 1 being directly related to CAD and so actionable by the servicing physician, while type 2 PACs are related to patient safety failures and can be improved by process improvement. Additionally, analyzing potentially avoidable admissions helps focus strategies in reducing them.

Risk Adjustment:

Once we have the observed PAC rates, we risk-adjust them for patient factors such as patient demographics, comorbidities collected historically, and for severity of illness or procedure using subtypes collected from the index stay and / or look-back period. This helps adjust for factors outside the providers control and levels the playing field for provider performance comparisons.

Unit of Analysis:
The unit of analysis is the individual episode.

Dependent Variable:
The dependent variable is a dichotomous variable indicating whether an episode had one or more claims assigned as a PAC (=1) or not (=0).

Independent Variables:
A number of patient-related “risk factors” or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months.

This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_CAD_all_codes_risk_adjustment 06.30.15.xls

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., unstable angina). Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_CAD_all_codes_risk_adjustment 06.30.15.xls

As mentioned previously, to avoid creating perverse incentives all comorbidities and subtypes are identified prior to or at the very start of the episode. None are identified during the episode period.

Statistical Methods

We use logistic regression to model the probability of at least one PAC occurring during the episode. Only comorbidities and subtypes are included in the models as covariates if they are present in at least 10 episodes to prevent unstable coefficients. No further model building is conducted after the initial models are built. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but it does not make it a priority that all covariates in the model be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity of a episode condition, and lets each regression model determine for itself which of the factors are more significant for a specific episode. Non-significant covariates in episode models can not overly influence predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

When more than one line of business is included in the data, separate models are calculated for each sample (i.e., commercial, Medicaid etc.).

Provider Attribution and calculating PAC rates by provider:

Once episodes are constructed they are attributed to providers based on one of the various attribution rules. For CAD, episodes are attributed to the primary care physician, internist or cardiologist with the highest count of office visits.

Using the logistic regression technique described above, a model is developed that gives estimates for each risk factor and subtype for the patients in the population analyzed. These estimates are used to develop patient-level probabilities for the occurrence of PACs. The patient-level probability estimates are summed to construct aggregated measures (e.g., facility/provider-level). This method is similar to the methods employed by the Centers for Medicare and Medicaid Services (CMS) and endorsed by the National Quality Forum (NQF) to construct similar facility- and practice-level measures (i.e., mortality, readmissions, etc.):

For each provider, the number of actual observed occurrences of the outcome is summed across all attributed patients with that episode, to give the observed PAC rates for the provider.

Similarly adjusted probabilities from the risk adjustment models are summed across all attributed patients to give expected PACs for the provider.

The observed sum is then divided by the summed probabilities (O/E). This number yields whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). This calculation yields a practice-level unstandardized performance ratio.

To facilitate accurate comparisons of rates across units of analysis, this ratio is then standardized to the community rate using the indirect method. Specifically, the provider-level rate is multiplied by the expected community rate, calculated as the sum of adjusted probabilities for every individual in the sample across all providers in the analysis. This measure, known as the standardized rate, represents what the unit's risk-adjusted rate would be for the outcome of interest if its patient population was reflective of the of the overall community.

The formula for this calculation is as follows:

Adj Outcome_j={(SUM Observed_ij)/(SUM Prob_ij)} × {(SUM Prob_i) / (# of episodes)}
Where individual is attributed to unit of analysis j (e.g., practice, provider, etc.)

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset will apply to another.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If patient related data is missing, the case is deleted from both the numerator and the denominator

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The information is based on a two-year claims database from a large regional commercial insurer. The database has over 3.2 million covered lives and $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

The methodology can be used on any claims database with at least two years of data and a minimum of 150 patients with the index condition or hospitalization. Having pharmacy data adds to the richness of the risk-adjustment models.
The calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.
The methodology has been tested on databases of several health plans as well as on a few employer databases.

No data collection instrument was used.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Team

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinician Office/Clinic, Hospital/Acute Care Facility, Other
If other: Across the care continuum

S.28. **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.) The individual complications are considered measurable components. Separate specifications are not required for this measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2740_CAD_Testing_Reliability_Validity_HCI3-635719664690615149.docx

NATIONAL QUALITY FORUM—Composite Measure Testing (subcriteria 2a2, 2b2-2b7, 2d)

Composite Measure Title: 2740
Measure Title: Proportion of Patients with coronary artery disease (CAD) that have a Potentially Avoidable Complication (during the episode time window)
Date of Submission: 06/30/15
Composite Construction:
☐Two or more individual performance measure scores combined into one score
☐ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)
☐ Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.
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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. 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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The information is based on a two-year administrative claims database from a large regional commercial insurer. The database contains medical and pharmacy claims on over 3.2 million covered lives and more than $25.9 billion in "allowed amounts" for costs.

1.3. What are the dates of the data used in testing? April 1, 2012 – December 17, 2014

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.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency

☐ health plan	☐ health plan
☐ other: Integrated Delivery System	☐ other:

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

There were a total of 5,840 providers in the data set. Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the reliability calculations. After this exclusion, there were 468 providers left.

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

After exclusions (see 2b.3.1 below), there were a total of 31,093 episodes of CAD were included in the testing and analysis. Patients in these episodes were, on average, 56.4 years of age (range 18-64) and 27% were female. We did not have race information on these patients. All patients for this analysis had a trigger inpatient claim of CAD as identified in our code tables.

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.

For the reliability analysis, we restricted the data to only providers with at least 10 attributed episodes. For risk adjustment, all episodes were used in the analysis, regardless of the provider to which they were attributed.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
None of the analyses included SDS variables.

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*)

We assessed the reliability of the measure to demonstrate that it sufficiently differentiates performance between providers using the beta-binomial method, which is applicable to measures of this type. Reliability is a measure that distinguishes the signal (the extent of performance variation between entities that is due to true differences

in performance) from statistical noise. Our approach follows directly from the methods outlined in the technical report "The Reliability of Provider Profiling: A Tutorial" by J.L. Adams.

Reference:

Adams JL. The Reliability of Provider Profiling: A Tutorial. Rand Corporation. http://www.rand.org/pubs/technical_reports/TR653.html.

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*)

The table below provides a summary of the reliability score for different minimum sample size thresholds. For complete results, refer to the workbook entitled, NQF_CAD_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab to see provider-specific results.

Reliability Scores	**Minimum # Episodes Per Provider**		
	>=10	**>=25**	**>=50**
# of Providers (%)	468 (100)	171 (37)	80 (17)
Median (IQR)	0.73 (0.61,0.83)	0.85 (0.79,0.91)	0.92 (0.88,0.95)
Range	0.50-1.00	0.72-0.99	0.84-0.99

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

Reliability scores can vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or variation across patients within providers) whereas a reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (see Adams, 2009 cited above).

Although scores among providers with at least 10 episodes were low, many had scores that met or exceeded the minimum acceptable level for reliability. Moreover, limiting providers to those with at least 25 or 50 episodes, scores were consistently good. These results demonstrate that the measure sufficiently differentiates providers' performance.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b2. VALIDITY TESTING

Note: *Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include*

assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.

2b2.1. What level of validity testing was conducted?

☐ **Composite performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

☐ **Systematic assessment of content validity**

☐ **Validity testing for component measures** *(check all that apply)*

***Note**: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☐ **Endorsed (or submitted) as individual performance measures**

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

☐ **Empirical validity testing of the component measure score(s)**

☐ **Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. 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)*

Content validity was built into the development of the definitions of potentially avoidable complications (PACs). This involved working with clinicians who are experts in their respective fields and specific to the episodes for which PACs are being measured. In particular, the clinical experts focused on whether or not a potentially avoidable complication can be deemed as such for a specific episode of care, and help defined and review all of the diagnosis and procedure codes for each PAC. The enclosed link lists clinicians who have participated in the various Clinical Working Groups (http://www.hci3.org/content/clinical-working-group-contributors). Some of the clinical experts have also participated in monthly webinars that highlight the clinical aspects of these measures (http://www.hci3.org/content/using-ecrs-providers).

In addition, we illustrate that our measure has face validity in several ways.

Beyond the up front work performed by clinical experts, the validity of the measure has also been tested in various real world settings. For example, we have presented results of claims data analyses that reveal the frequency and costs of PACs to physicians in several different healthcare systems involved in our pilot site implementations, as well as to medical directors from the employer coalitions and the health plans that provided the dataset to run the analyses. Some of these implementations include the Pennsylvania Employee Benefits Trust Fund and local provider groups and hospital, Horizon Blue Cross Blue Shield of NJ and many physicians and health systems.

In addition, we have performed dozens of analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:

- Vermont Payment Reform Commission
- Maine Health Management Coalition
- WellPoint / Anthem CT

- NY State Medicaid
- CT Medicaid
- CO All-payer Claims Database, Center for Improving Value in Health Care

These analyses and their results have influenced, and continue to influence, the development of various public reporting, payment reform and delivery system reform efforts. To-date, we have never experienced either wholesale or partial rejection of the results of analyses showing rates of PACs, which demonstrates the level of acceptability – face validity – of the measures from the payer, policymaker, employer and payer communities.

As importantly, measures of potentially avoidable complications have face-validity with consumers. In a series of focus groups, Judy Hibbard and colleagues[1] examined the impact of presenting information about price and quality of certain providers in influencing the decisions of consumers. They tested the validity of PACs as a discriminator of quality, as well as other measures of quality, and used the dollar symbol to illustrate the level of price, much like is done for restaurant reviews. When the PAC measure was used, respondents selected the providers with the lowest PAC rates with a high level of confidence in choice, and used it as a surrogate for a strong quality signal. To the contrary, when more standard measures of quality were used, consumers tended to ignore them and use price as a surrogate for quality. As such, what the researchers found is that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Finally, our measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs)[2], Hospital Inpatient Quality Reporting measures [3], Avoidable Readmissions [4,5], AHRQ defined patient safety indicators (PSIs) [6], NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates [7].

References:

[1] Hibbard JH, Greene J, Sofaer S, Fiminger K, and Hirsh J. An Experiment shows that a well-designed report on Costs and Quality can help consumers choose High-Value Health Care. *Health Affairs* 2012; 31(3): 560-568. doi: 10.1377/hlthaff.2011.1168

[2] CMS defined Hospital Acquired Conditions: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

[3] CMS operated Hospital Inpatient Quality Reporting Program: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

[4] Jencks SF, Williams MV, and Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. N Engl J Med 2009 (Apr); 360 (14): 1418-1428. doi: 10.1056/NEJMsa0803563.

[5] Casalino LP, Pesko MF, Ryan AM et.al. Small Primary Care Physician Practices have low rates of Preventable Hospital Admissions. Health Affairs, 2014; 33(9): 1-9. doi: 10.1377/hlthaff.2014.0434.

[6] Agency of Healthcare and Quality defined Patient Safety indicators: http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx

[7] NQF endorsed measures: Quality Positioning System: http://bit.ly/1E5ZdP7

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Not applicable.

2b2.4. 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?*)

Given the significant clinical input that went into developing the measure, the widespread use and acceptance the measure has gained among a wide variety of individuals and organizations across the health system (public and private payers, clinicians, consultants, patients, etc.) [1-13], and the parallels between this measure and other measures that are in widespread use, this demonstrates that the measure has strong face validity.

References:

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)

2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.

3. de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.

4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.

5. Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.

6. Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.

7. François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)

8. de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

9. Rastogi A, Mohr BA, Williams JO, Soobader MJ, de Brantes F. Prometheus Payment Model: Application to Hip and Knee Replacement Surgery. Clin Orthop Relat Res 2009; 467(10): 2587-2597.

10. de Brantes F and Rastogi A. Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. The Commonwealth Fund 40, publ. 2008; 1146:1-14.

11. de Brantes F, Gosfield A, Emery D, Rastogi A and G. D'Andrea, "Sustaining the Medical Home: How Prometheus Payment Can Revitalize Primary Care", Robert Wood Johnson Foundation Report, May 2009, http://www.rwjf.org/pr/product.jsp?id=42555, accessed October 2009.

12. de Brantes F, Camillus J. Evidence-informed case rates: a new health care payment model [Internet]. New York (NY): Commonwealth Fund; 2007 Apr [cited 2007 May 20]. Available from: http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=478278, Accessed Aug 1 2013.

13. Satin DJ, and Miles J. Performance Based Bundled Payments: Potential Benefits and Burdens. Available from: http://student.med.umn.edu/p4p-new/sites/default/files/MN%20Med%20Bundles%20Special%20Report%20-%20Satin.pdf, Accessed Aug 1 2013.

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

No formal exclusion testing was done since no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers.

Exclusions include exclusions of "patients" as well as "claims" not relevant to CAD care. Please refer to the enclosed excel workbook entitled (NQF_CAD_all_codes_risk_adjustment_06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
d. If the patient does not have at least 12 months of claims in the database (this helps eliminate incomplete episodes).
e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure if they meet one of the following criteria:
a. If none of the diagnosis codes on the claim are on the list of "triggers" or relevant diagnosis codes (either typical Dx or PAC Dx) for CAD.
b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for CAD.
c. If the CAD trigger hospitalization also triggers a major surgical procedure such as coronary bypass procedure or angioplasty, suggesting that CAD may be a comorbidity or an indication for the surgery.
d. The "principal" diagnosis on an inpatient stay claim during the episode time window triggers its own episode
e. The procedure code on a claim during the episode time window triggers its own episode

2b3.2. 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*)

We started with a total CAD population of 63,972 episodes. After all the exclusions were applied, the remaining CAD population included in the analysis consisted of 34,016 episodes. As mentioned above, no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers. As such, no formal exclusion testing was done.

2b3.3. 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*)

No formal analysis was done on the impact of exclusions on performance scores.

Descriptive Explanation:

Exclusions of patients were for the following reasons. Some are for comparative purposes and some for medical reasons.

(a) Comparative Purposes:
We excluded patients that did not have complete enrollment for the entire episode time window. This was done to ensure that the database had complete information on patients to be able to create the entire episode. Including patients with only a partial episode window could distort the measure by artificially reducing the actual count of patients with PACs.

(b) Medical Reasons:
Patients with outlier costs (less than 1st percentile value or greater than 99th percentile) were considered to be different from the general pool, and excluded from both the numerator and the denominator. This is another way to ensure that episodes are complete (because incomplete episodes may have very low costs), and do not bring in random noise into the analysis due to inappropriate codes or services (high outliers).

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1./S13 What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model** with 170 potential risk factors and episode specific subtypes
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.1.2. If an outcome or resource use 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.

2b4.2/S14. Identify the statistical risk model variables (Name the statistical method – e.g., logistic regression and list all the risk factor variables.
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_CAD_all_codes_risk_adjustment_06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., Previous CABG, PCI). Subtypes are specific to each unique episode and are included in the models only if they are present at the start of the episode. Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_CAD_all_codes_risk_adjustment_06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Candidate comorbidities and subtypes were included in the models as covariates if they were present in at least 10 episodes to prevent unstable coefficients.

The detailed list of all possible risk factors and subtypes is provided in the tables below.

2b4.2.1/S15. Detailed risk model specifications including coefficients, equations, codes with descriptors, definitions(may be attached in an Excel or cvs file)
All Risk Factors with their coefficients are detailed in the enclosed workbook called NQF_CAD_all_codes_risk_adjustment_06.30.15.xls – Please reference the tabs titled Risk Factor Prevalence and Risk Model.

2b4.3. Describe the conceptual/clinical <u>and</u> statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Risk factors are comorbidity indicators collected from historical claims before the start of an episode. These are universally applied across all episodes. This list was selected based on input from clinical experts in clinical working groups. In addition, the Clinical Working Groups identified episode specific severity markers that were called episode subtypes and they help distinguish an episode as being more severe than another.

All risk factors and subtypes must be present prior to, or at the start of the episode and are identified using diagnosis codes in the patient's historical claims.

To be included in the risk adjustment models, any risk factor or subtype must be present in at least 10 episodes. Beyond this no further model building was conducted to add or remove risk factors or subtypes from the model after it was initially run. This reflects a desire to explain as much variation in the probability of having a PAC

as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates cannot overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

As explained above, no formal analysis was conducted to select risk factors. In fact, all potential risk factors and subtypes with a count of at least 10 episodes were retained to serve as predictors. The goal was to achieve a more complete explanatory model rather than achieve parsimony.

Please reference the tabs titled Risk Model in the NQF_CAD_all_codes_risk_adjustment_06.30.15.xls workbook to see the list of risk factors that met the selection criteria.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Not Applicable since our analysis did include SDS variables

2b4.5. 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*)

Model Development Approach

We used logistic regression to model the probability of at least one PAC occurring during the episode. The model included all covariates that were identified through the process above. No further model building was conducted after the initial model was run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates can not overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

For a more complete description of the risk adjustment approach, please see the document entitled, "PACs and Severity Adjustment Fact Sheet" that accompanies this submission.

Approach to Model Testing and Validation

To determine the validity and performance of the model, we used the split sample method to divide the patient sample randomly into: 1) the model building data set (80% of the sample) and 2) the test data set (20% of sample. The model was built using logistic regression on the first data set and then the coefficients from the development model were tested in the second dataset. Area under the curve (AUC) and the c-statistic were used to compare the predictive ability of the model in each of the data sets. Hosmer-Lemeshow Goodness-of-Fit tests and comparisons of observed to expected probabilities across risk deciles were further examined to assess the model's overall predictive accuracy.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

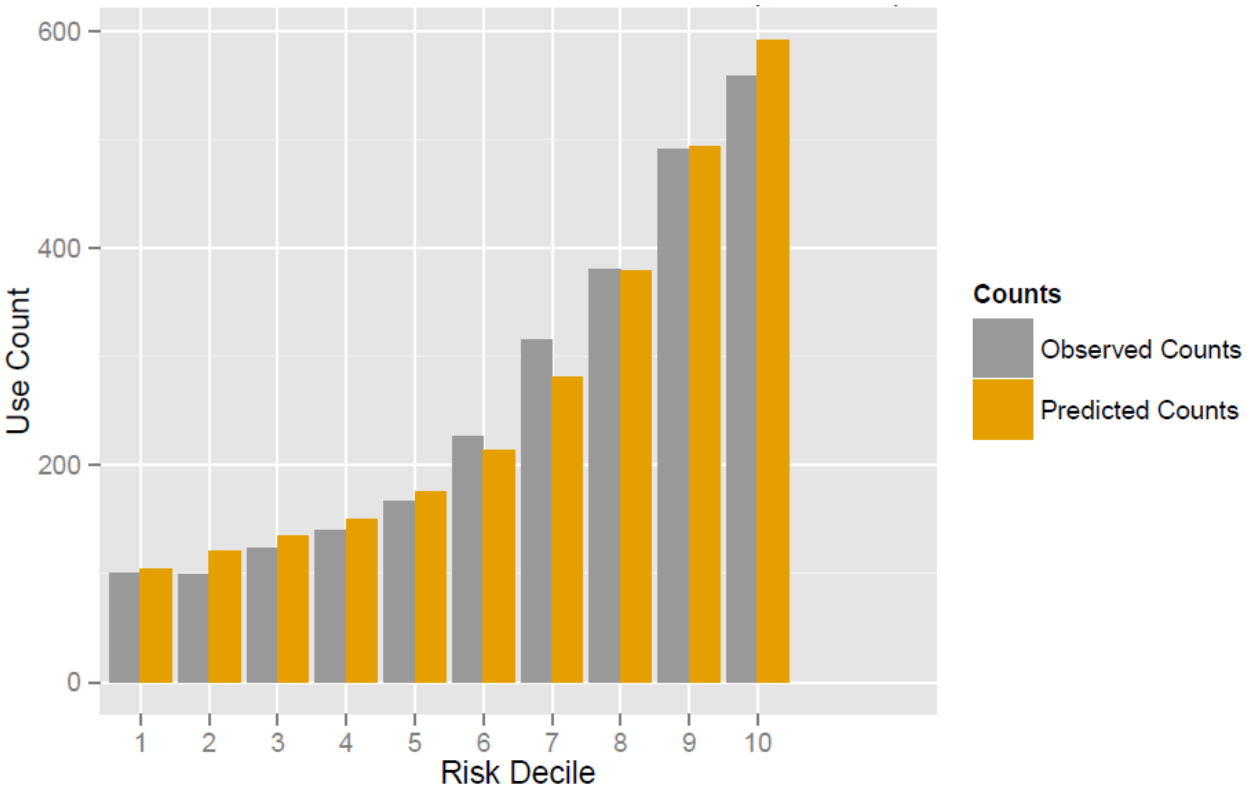
Sample	Accuracy (%)*	AUC
Test	75.4%	0.803
Validation	74.5%	0.792

*Episodes with predicted probabilities <50% were classified as having a predicted 0 (not having a PAC). Episodes with predicted probabilities >50% were classified as having a predicted 1 (having a PAC)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Sample	Chi Square	Degrees of Freedom	p-value
Test	199.5	8	<0.0001
Validation	51.8	8	<0.0001

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:



2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The C statistic is a measure of the extent to which a statistical model is able to discriminate between a patient with and without an outcome. The c-statistic ranges from 0.5 to 1.0. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong [1]. The purpose of the model is to adjust for patient-related factors. The remaining unexplained differences in PAC rates are due to factors that could be controlled by all providers that are managing or co-managing the patient, during the entire episode time window.

The c-statistics of the testing and validation samples (0.803 and 0.792, respectively) indicate that the risk models have strong discriminatory power. Indeed, the accuracy values show that the model correctly predicts whether an episode had or did not have a PAC about 75% of the time, well above what would be expected if the predictions were made at random (i.e., 50%). Although the H-L test was significant for the testing sample, meaning that the model is not a good fit, this test is generally known to be sensitive to the number of groupings used and sample sizes. Additionally, the risk decile plot shows that the model predicts PACs similarly to the observed PACs across all deciles.

Overall, the results strongly suggest that the models have strong predictive power.

Reference:
[1] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Edition)*. New York, NY: John Wiley & Sons; 2000.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

NA

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To directly compare PAC rates across providers or providers while also appropriately accounting for differences in patient severity, we calculated a risk-standardized PAC rate for each provider. This method is similar to calculations used by others for reporting outcomes measures [1]. For each provider, the ratio of observed attributed episodes with PACs to the expected number of attributed episodes with PACs given the patient's risk factor and estimated from the risk-adjustment model was calculated. This number yielded whether the provider had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). We then multiplied this ratio by the overall expected PAC rate across all providers to obtain the risk-standardized PAC rate for the provider. This measure represents what a provider's PAC rate would be if its patient population was reflective of the overall population.

Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. Comparison of risk-adjusted PAC rates gives a measure of the provider's relative performance. Our analysis compared risk-standardized PAC rates across providers. We analyzed various descriptive statistics including the range in PAC rates, medians, interquartile range, etc.

Reference:

[1] See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

2b5.2. 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*)

Summary of Unadjusted and Adjusted Performance Scores Across Providers:

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25
Unadjusted		
Median (IQR)	40% (30%, 55%)	39% (32%, 51%)

Range	0%-100%	8%-94%
Adjusted (RSPR)*		
Median (IQR)	40% (33%, 47%)	41% (34%, 46%)
Range	0%-84%	11%-71%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_CAD_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
The variation in risk-adjusted rates suggests there are meaningful differences in performance between providers in risk-standardized PAC rates for patients with an episode of CAD.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This criterion is directed 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 SDS 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.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not performed

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

If patient related data is missing, the entire patient is excluded from the numerator as well as the denominator.

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence.

According to our measure definition, in constructing the measure it is possible for a provider to have only one or some types of PACs and not others. Alternatively, the provider may have all PAC types occur for their patients. The measure only considers whether any PAC occurred regardless of the type, and all PAC types are weighted equally, therefore there is no potential for the absence of specific PAC types to bias performance scores for individual providers.

For these reasons, no formal analyses were done on missing data.

2b7.2. 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*)

Not applicable

2b7.3. 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*)

Not applicable

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook entitled NQF_CAD_all_codes_risk_adjustment_06.30.15.xls provides outputs from empirical analysis. The tab labeled "PAC overview" demonstrates percentage of episodes that had at least one PAC, and provides the breakdown of PACs: 1) by the type of PAC whether directly related to index condition or due to patient safety failures; 2) the setting of the PAC, whether seen in the in-patient setting, out-patient facility or during professional visits; and 3) preventable hospitalizations.

The "PAC Drill Down Graph" provides further detail on each component of the PAC and their frequency. As can be seen by the individual counts and the graph, while each individual PAC may have such small occurrences that no meaningful comparisons in provider performances could be made; together, they add value to provide a comprehensive picture that result in meaningful numbers. The aggregation of PACs to a comprehensive, composite measure, in itself provides the parsimony that is so desirable.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

All PACs, as clinically defined by the subject matter experts were used with equal weighting. Since the emphasis of the PAC measure is to identify the occurrence of PACs in any setting, a simple and straightforward approach was adopted.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc.*; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

No formal analysis was performed.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Additionally, if a patient had one or more PACs, it is simply counted as a 1, i.e., flagged as having a PAC. The measure only considers whether any PAC occurred regardless of the type, or the site, and all PAC types are weighted equally. Therefore, no formal analysis of individual components was performed.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

We chose not to weight the components of the measure.

Considerations were given to the fact that preventable hospitalizations may be given more weight, than PACs identified in a doctor's office. Similarly PACs in an in-patient setting may have more serious implications on a patient's ultimate outcome, than PACs occurring in an outpatient setting. Additionally, preventable hospitalizations as well as index hospitalizations, each with longer lengths of stay, may have serious PACs. But how do we weigh these effects? An alternative model was considered, where cost could be considered as a surrogate for the weights. Higher cost PACs could imply more serious PACs. However, differences in costs could be driven by many issues other than the PAC itself, such as unit price of the service, method of reimbursements, contracting arrangements etc.

Furthermore, in-patient facility billing does not allow for the distinction of PAC related costs from other costs within the stay. We would fail to capture PAC related costs within the stay and potentially underweight those. As a result, the decision was made to avoid weighting and keep the measure as a straightforward count.

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

Measuring all providers with the same yardstick will provide consistent results and reasonable comparisons over time. If the goal is to reduce PACs, then the PAC measure as was constructed with the help of various experts in the field would provide reasonable comparisons. A word of caution however pertains to the sample size of the provider panel before making any reasonable conclusions.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved for high reliability in one dataset would apply to another.

2d3. Empirical analysis demonstrating that the approach for handling missing data minimizes bias (*i.e., achieves scores that are an accurate reflection of quality*)**.**

Note: *Applies to the overall composite measure; the focus is on missing data rather than exclusions, which are considered in 2b3.*

Please refer to section 2b7

2d3.1. What is the overall frequency of missing data and the distribution of missing data across providers?

2d3.2. Describe the method used to compare approaches for handling missing data (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2d3.3. What were the statistical results obtained from the analysis of missing data? (e.g., *results of sensitivity analysis of effect of various rules for missing data; if no empirical analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

2d3.4. What is your interpretation of the results in terms of demonstrating that the approach used for missing data minimizes bias? (i.*e., what do the results mean in terms of supporting the selected approach for missing data; if no empirical analysis, provide rationale for the selected approach for missing data*)

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

As part of our general implementation of these measures and related analyses, we have worked through dozens of different and sometimes very large datasets. From Medicare to Medicaid to regional and national commercial carriers, as well as individual employers, the principal lesson learned is the heterogeneity of the data sets and the significant variability in fill rate of critical data elements. As a result, we have created highly specific recommendations for which data elements are required to ensure measure validity, the accuracy of those data elements, and their completeness in the dataset. When claims datasets are organized in the way we specify in the measure analysis, and contain the coding information required, the analysis of the measure and its results are highly reliable.

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 calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.

We also plan on providing a limited automated analysis, at no cost, on our website.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Payment Program Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/ Blue Cross Blue Shield of New Jersey http://www.horizonblue.com/ Pennsylvania Employee Benefits Trust Fund https://www.pebtf.org/ Quality Improvement (Internal to the specific organization) Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

4a.1. For each CURRENT use, checked above, provide:

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

Measures associated to potentially avoidable complications (PACs) are in use today with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system (public and private payers, clinicians, consultants, all-payer claims database stewards, etc.) [1-8]. They are being used in various capacities in different pilot site implementations. To name a few:

•BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
•BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers

In addition, the PAC measures are incorporated by the following organizations in their bundled payment programs:

•BCBSSC – for CABG and PCI programs
•Horizon BCBSNJ– for CHF and CABG programs
•BCBSNC
•PEBTF in PA

http://www.ajmc.com/interviews/Lili-Brillstein-on-How-Bundled-Payments-Are-Tranforming-Healthcare

In these programs they look at PACs related to the measure for process improvement activities and for practice re-engineering.

We have created reports for rates of PACs for the following organizations:
-Vermont Payment Reform
-Maine Health Management Coalition

-WellPoint / Anthem CT
-NY State Medicaid
-CT Medicaid
-CO All-payer Claims Database, Center for Improving Value in Health Care

There are several companies that are leveraging these measures to create analytics and software for customers – these include HealthQx, Aver Informatics, McKesson, and TriZetto.

Below are some references that highlight our work with Potentially Avoidable Complications (PACs).

1.Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)
2.Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.
3.de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.
4.de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.
5.Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.
6.Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.
7.François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)
8.de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

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

4a.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.*)
Measures associated with PACs are currently in use as described in the prior section. In addition, we are working with several not-for-profit and for-profit organizations to provide them with the algorithms needed to calculate rates of potentially avoidable complications. Some of these organizations include:

Fair Health – based in NY and whose mission is to increase transparency of provider cost and quality,

CastLight – based in CA and serving large employers. We currently provide CastLight with Bridges To Excellence recognitions and will work with them to augment provider transparency by using PAC measures,

MA APCD (Massachusetts All Payers Claims Database) Council – we currently have an agreement in place with the MA APCD Council to produce PAC measures on hospitals and physicians and report back to the council with tests of reliability and validity of the measures. The purpose is to authorize the publication of these measures,
Maryland Health Care Cost Commission – we have a two year agreement to produce measures of cost and quality for public dissemination.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

We do not have any public information to share about the improvements in rates of potentially avoidable complications, as the implementation of these measures is too recent to provide valid comparisons. Further, some of the definitions of PACs have changed since the measures were initially endorsed, making comparisons even more difficult and unreliable.

Nevertheless, the variation in performance scores presented in Section 1b.2 indicates that there are differences between providers in their risk-adjusted PAC rates (higher scores equal worse performance). This suggests that real opportunities exist to identify lower performing providers and reduce the overall occurrence of PACs.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Performance results provide summary PACs rates by provider, which can be used by payers and providers in a number of ways to improve the quality of care.

From the payer perspective, payers can use this information to 1) create a high-value provider networks, 2) work with high-value providers to share best practices, 3) incentivize low-value providers to improve, 4) modify their insurance design to activate consumers to select the right care from the right providers at the right time.

From the provider perspective, providers can 1) view services and activity for their patients longitudinally across the entire care continuum, such as frequency of readmissions and ED visits and drill down on patients with high PAC rates, 2) review actionable drill down reports to identify the most frequent PACs across all patients to create care pathways and process improvement plans to impact the most frequent PACs.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended consequences were reported, but there is the potential for:
1. Under-coding of PACs in the claim stream resulting in under-reporting the actual rate and/or providers gaming the measures
2. Payers calculating the measures even with inadequate sample sizes and using the results to penalize providers

The measure is designed for transparency efforts and to spur quality improvement. Detailed PAC reports can help providers identify areas of quality improvement. Even detailed reports of small samples of patients can be helpful for quality improvement purposes, but not for public reporting. To mitigate the potential for invalid provider comparisons, we specify in this submission the minimum sample size needed to ensure the reliability of a provider's score. Ultimately, there isn't any good way to prevent provider gaming of the measure by under-coding claims, however, under the current DRG payment methodology, many providers would be penalized by under-coding PACs since these codes often result in the assignment of more complicated DRGs.

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)**

0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.

5a. Harmonization

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 completely harmonized?

No

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

Some of the measures listed in the prior section are, fully harmonized with the submitted measure, in particular, 0705, 0708, and 0709. Other measures such as 0337 and 0450 are in fact, subsets of our measure. However, there are some measures that are not harmonized, in particular the 30-day all-cause readmission measures and the Hospital wide all-cause readmission measure. While the submitted PAC measure include hospitalizations and readmissions that occur during the episode time window, the hospitalizations, by definition, have to be relevant to the underlying condition. For chronic conditions, most relevant hospitalizations within the entire episode time window are considered potentially avoidable. PACs include readmissions and are designed to enable accountability at the locus of provider control as well as some shared accountability between settings, centered around a patient, and for a specific medical episode of care. In that sense, they are consistent with the all-cause 30-day readmission rates, but represent a subset of those admissions. However, they do extend to the entire episode time window. As such, the PAC measures, as submitted, don't create added burden of reporting because the readmissions reported are simply a part of the broader 30-day all-cause readmission measures already endorsed by NQF. Because PAC measures are comprehensive, they include patient safety events as well as other adverse events, including hospitalizations and ED visits during the entire continuum of care. As a result, they are a comprehensive measure of avoidable complications for a specific medical episode. The data collection for all of the HCI3 measures is automated by a software package and is fully harmonized with all other PAC measures. A single download automates creation of all reports related to each of the PAC measures.

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.)

PAC measures are composite measures representing "all-cause harms". They look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. . Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence. PACs look at readmissions, emergency room visits, adverse events due to errors of omission or commission. They look at complications that are due to patient safety failures, and also those directly related to the index condition. These are all a cause of significant waste and quality concerns. As such, the measure can provide clinicians with an overall and comprehensive view, in one measure, of all potentially avoidable complications for a patient and drive quality improvement efforts.

For clinicians and facilities increasingly engaged in value-based payment efforts and/or driving quality improvement for population health, the value of a PAC measure over a series of related, but more discrete measures, is that one can better determine if the sources of complications primarily stem from activities within the facility or outside the facility, and the specific nature of the complications that have a higher frequency of occurrence. While individual components of the PAC measure may have small frequencies and may be difficult to interpret with regards to provider performance or actionability, aggregating all the PACs into a comprehensive, composite measure provides the parsimony that is so desirable. For providers, it's far easier to construct a quality dashboard from a parsimonious set of measures, and that's what PAC measures offer.

Further, as a comprehensive outcome measure, PACs are also useful for public transparency of quality, as substantiated by the research from Judy Hibbard and colleagues previously cited in the "testing" section of this submission. As a comprehensive outcome measure, they are easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PACs_and_Severity_Adjustment_Fact_Sheet_HCI3-635719632690865141.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.2 **Point of Contact:** Francois, de Brantes, Francois.debrantes@hci3.org, 203-270-2906-
Co.3 **Measure Developer if different from Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.4 **Point of Contact:** Amita, Rastogi, Amita.rastogi@hci3.org, 213-934-9624-

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.
From 2006 onwards, and under the auspices of various funding organizations, HCI3 has convened and managed, or helped to convene and manage, Clinical Working Groups to inform the development and refinement of the measures. For example, in 2011, 2012 and 2013, HCI3 worked collaboratively with the American Board of Medical Specialties and the American Medical Association's Physicians Consortium for Performance Improvement, under a federal contract, to convene and get input from various clinical experts on definitions of episodes of care and their sequelae, including avoidable complications.

Some of the clinical experts that have contributed to our work include:
-Dr. John Allen, American Gastroenterology Association (AGA)
-Dr. Morton Arnsdorf, Cardiologist, University of Chicago, IL

-Dr. Peter Bach, Memorial Sloan Kettering Cancer Center (MSKCC)
-Dr. Peter Basch, Primary Care, Medstar Health, DC
-Dr. Justin Beckelman, Radiation Oncology, University of Pennsylvania, PA
-Dr. Debra Bingham, Executive Director, California Maternal Quality Care Collaborative (CMQCC) at Stanford University, CA
-Dr. John Birkmeyer, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Linda Bosserman, Wilshire Oncology Medical Group, CA
-Dr. Matthew Brengman, American Society of Metabolic and Bariatric Surgery (ASBMS)
-Dr. Joel Brill, American Gastroenterology Association (AGA)
-Dr. George Cautilli, Cautilli Orthopedic Surgical Specialists PC, Yardley, PA
-Dr. Ashwini Davison, Internist, Johns Hopkins Hospital, MD
-Dr. James Denneny, III, American Academy of Otolaryngology – Head and Neck Surgery (AAO-HNS)
-Dr. Chris Gallagher, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Robert Haralson, III, American Academy of Orthopedic Surgeons (AAOS)
-Ms. Dawn Holcombe, Executive Director, Connecticut Oncology Association, CT
-Dr. Colin Howden, American Gastroenterology Association (AGA)
-Dr. John Knightly, American Association of Neurological Surgeons (AANS)
-Dr. Larry Kosinski, American Gastroenterology Association (AGA)
-Dr. Nalini Krishnan, Obstetrics & Gynecology, MN
-Dr. Kelly Kyanko, Internist, NYU School of Medicine, NY
-Dr. Tara Lagu, Internist & Infectious Disease, Baystate Medical Center, MA
-Dr. Robert Lee, Society of Thoracic Surgeons (STS)
-Dr. Alex Little, Society of Thoracic Surgeons (STS)
-Dr. Michael London, Orthopedic Surgeon, OMNI Orthopedics, OH
-Dr. Elliott Main, Obstetrics & Gynecology, California Pacific Medical Center, CA
-Dr. Constantine Mantz, 21st Century Oncology, FL
-Dr. Joseph Messer, Cardiologist, Rush University Medical Center, IL
-Dr. David Metz, American Gastroenterology Association (AGA)
-Dr. Ronald Nahass, Infectious Disease Care, NJ
-Dr. Ajay Nehra, Urologist, Rush University Medical Center, IL
-Dr. Francis Nichols, Society of Thoracic Surgeons (STS)
-Dr. Patrick O'Connor, Primary Care, HealthPartners, MN
-Dr. Sara Perkel, National Comprehensive Cancer Network, PA
-Dr. David Peura, American Gastroenterology Association (AGA)
-Dr. John Ratliff, American Association of Neurological Surgeons (AANS)
-Dr. Steven Schutzer, Connecticut Joint Replacement Institute, CT
-Dr. Leif Solberg, Primary Care, HealthPartners, MN
-Dr. Scott Sporer, Midwest Orthopedics at Rush, Chicago IL
-Dr. Bonnie Weiner, Cardiologist, Worcester Medical Center, MA
-Dr. Jonathan Weiner, Bariatric Surgery codes, Prof of Health Policy and Management, Johns Hopkins University, MD
-Dr. Janet Wright, Cardiologist, Northstate Cardiology Consultants, CA

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** Yearly
Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

#2747 Proportion of Patients with Heart Failure (HF) that have a Potentially Avoidable Complication (during the episode time window), Last Updated: Aug 18, 2015

NATIONAL QUALITY FORUM

Measure Information - Composite

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 subcriterion 1b).

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

Brief Measure Information

NQF #: 2747
Measure Title: Proportion of Patients with Heart Failure (HF) that have a Potentially Avoidable Complication (during the episode time window)
Measure Steward: Health Care Incentives Improvement Institute Inc. (HCI3)
Brief Description of Measure: Percent of adult population aged 18 + years who triggered an episode of heart failure (HF), are followed for at least one-year, and have one or more potentially avoidable complications (PACs). PACs may occur any time during the episode time window. Please reference attached document labeled NQF_HF_all_codes_risk_adjustment_06.30.15.xls, in the tabs labeled PACs I-9 and PAC I-10 for a list of code definitions of PACs relevant to HF.
We define PACs as one of two types:
(1) Type 1 PACs - PACs directly related to the index condition: Patients are considered to have a PAC, if they receive services during the episode time window for any of the complications directly related to HF, such as for hypotension, acute heart failure, fluid and electrolyte disturbances etc.
(2) Type 2 PACs - PACs suggesting Patient Safety Failures: Patients are also considered to have a PAC, if they receive services during the episode time window for any of the complications related to patient safety failures such as for sepsis, infections, phlebitis, deep vein thrombosis, pressure sores etc.
All relevant admissions in a patient with HF are considered potentially avoidable and flagged as PACs.
PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook labeled NQF_HF_all_codes_risk_adjustment_06.30.15.xls serves as an example. The tab labeled PAC overview gives the percent of HF episodes that have a PAC and the tab labeled "PAC drill down" gives the types of PACs and their frequencies in HF episodes within this dataset.
The information is based on a two-year claims database from a large regional commercial insurer. The database had over 3.2 million covered lives and over $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.
Developer Rationale: Each individual PAC, when measured in isolation, provides a very limited picture of the performance of the provider(s) who are managing or co-managing the care of the patient. However, looking at all the PACs that may occur individually or concurrently in a patient with a given episode provides a comprehensive picture of the care received by the patient for that particular condition or illness.

Additionally, the frequency of occurrence of individual PACs may be so low that it may require very high sample sizes from individual providers to achieve any meaningful and reliable comparisons. But aggregating all the PACs into a single quality metric creates meaningful scores that can be compared across providers even with relatively smaller sample sizes.

Additionally, a comprehensive measure is easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

Numerator Statement: Outcome: Number of patients who triggered an episode of heart failure (HF), are followed for at least one-year, and had one or more potentially avoidable complications (PACs) during the episode time window.

Denominator Statement: Adult patients aged 18 years and above who triggered an episode of heart failure (HF) and are followed for at least one-year.
Denominator Exclusions: Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the HF measure if they are considered not relevant to HF care.

Measure Type: Outcome
Data Source: Administrative claims
Level of Analysis: Clinician : Group/Practice, Clinician : Team
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

1d.1. **Composite Measure Construction:** any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)
Component Measures (if endorsed or submitted for endorsement):

Is this a MAINTENANCE measure submission? ☐ **Yes** ☒ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: n/a **Most Recent Endorsement Date:** n/a

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health *outcomes* measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if health outcomes measures agree the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

The developer provides the following evidence for this outcome measure:

- This new risk-adjusted (by age, gender and clinical co-morbidities) outcomes measure that assesses the proportion of adult patients with claims triggered Heart Failure (HF) with at least one Potentially Avoidable Complications (PAC) within 12 months of HF triggered claims data. Based on NQF's criteria, this measure is considered an "any or none" composite measure that assesses if 1 or more PACs or "care defects" have occurred for the index episode. For this composite measure, the individual complications considered the measurable components. PACs are classified in two types: 1) related to CAD, and 2) related to Patient Safety Failures.
- PACs are classified in two types: 1) related to HF, and 2) related to Patient Safety Failures, combining the 2 types into a single PAC rate to calculate the proportion of patients with 1 or more PAC. PACs are considered unwarranted health outcomes that combine concepts from AHRQ PSIs, PQIs and the CMS HACs, and episode-specific PACs into all-cause patient harms that are measured during an index condition for use at the practice, medical group, provider system or purchaser/payer levels to identify quality of care gaps between practices and hospitals.
- The developer links primary & secondary prevention care gaps, poor patient education, poor care coordination and

poor follow-up increase unnecessary ER visits, hospitalizations, readmissions, and mortalities to increased PACs, and state that PACs for HF patients should occur rarely in well-managed patients.

- The evidence for Patient Safety Failure PACs is described to be within the influence of the measured entity, and does not describe the influence of non-healthcare-related impacts on PAC rates. The progression of the episode condition, illness or disease is also not mentioned as a contributor to PAC rates in the evidence.
- In addition to linking processes of care to outcomes, the developer provides an extensive PAC literature review in sections 1a.2. and 1a.2.1. for HF, Patient Safety Failures & processes of care, as well as background information on the process for PAC development.

Questions for the Committee:

- *Does sufficient evidence exist connecting Patient Safety Failures to the HF index episode?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*

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

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

- The developer provides CAD prevalence & impacts data, rationales and general information on PAC measure utility and applicable setting use.
- HF PAC performance gap data are calculated from PROMETHEUS administrative claims data from April 1, 2012 through December 17, 2014, for providers with ≥ 10 attributable index episodes. The data includes 81 of 2110 (3.8%) providers from 6,025 of 25,284 (23.8%) index episodes in 3,258,706 unique beneficiaries.

Unadjusted PAC Rates:		**Risk-Standardized PAC Rates (RSPR):**	
Median (IQR):	40.6% (30.8%, 57.1%)	Median (IQR):	39.9% (32.4%, 46.2%)
Range:	9.1% - 80%	Range:	14.5% - 67.9%

- Limited descriptive data on the patient, provider and payer are provided. The developer provides "Overview" and "Drill Down" PAC rates in the spreadsheet demonstrating gap, though PAC rates for individual complications are not provided.
- The developer does not provide data on disparities.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - This is an outcome measure - Potentially Avoidable Complication
   - Measure does include items from the AHRQ Patient Safety Indicators, PQIs and CMS Hospital Acquired

Conditions but also includes a myriad of other events/conditions that are claimed to be PACs related to HF. No distinction is made regarding the severity of a PAC (only HF-related or Patient Safety Failure). No distinction is made regarding the number of PACs that occur in a single patient. It is any or none.

- The developers define a very broad composite measure of potential avoidable complications (PACs) that lumps any heart-failure related events (hypotension, respiratory insufficiency, acute heart failure, pulmonary edema, fluid and electrolyte disturbances, etc) with any patient safety failures (sepsis, infections, phlebitis, deep vein thrombosis, pressure sores, etc).
- The assertion that well-managed patients with heart failure should rarely incur a potentially avoidable complication such as an emergency room visit or hospitalization related to heart failure is ludicrous.
- The developer cites data that support the efficacy of multi-pronged interventions for improving primary and secondary prevention care (patient education, care coordination and follow-up) with a reduction in ER visits, and all-cause hospitalizations, but not mortality.
- They cite no data to support the composite measure of PACs as a quality metric, but argue for it on face validity grounds.
- The developers cite a 42.5% rate of PACs in a large regional database, but an actual preventable hospitalization rate of only 5% of all HF episodes.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- There is a lengthy explanation of how failures in care can lead to PACs and readmissions. There is no direct cause and effect data evidence (for the myriad of PACs listed) -- just associations. There are approximately 846 different diagnosis codes identified as HF-related PACs. The PACs are derived from two years of claims data from a large regional commercial insurer. One PAC is listed as ""adverse effect of drug - correct use"" How is this avoidable?
- Also for the data to be reliable more than 10 attributable episodes per provider were required. Only 81 of 2110 (3.8%) met this criteria.
- The evidence connecting Patient Safety Failures to the HF index episode is lacking.
- I do not see a compelling argument for making an exception to the evidence criteria, nor do the developers cite any systematic assessment of expert opinion beyond the developers.

1b. ***Committee's Comments on Performance Gap:***

- Developer provides adjusted and unadjusted PAC occurance rates. There is no optimum rate/lowest rate provided. There is an assumption that PACs should never/rarely occur. That may be true for some, but not all of the PACs listed.
- Unadjusted PACs 40.6% (range 9.1-80%) Risk adjusted PACs 39.9 (range 14.5-67.9%)
- The developers cite data showing wide variance in current performance on this proposed measure (14.5-67.9% risk-standardized PAC rate). They cite no quality improvement metrics or goals for ""acceptable performance.""
- No data was provided on disparities in care for population subgroups. I seem to remember seeing data on racial disparities for HF care.
- I do not think this measure should be indicated as disparities sensitive.

1c. ***Committee's Comments on Composite Performance Measure:***

- This is a composite measure. If a patient as 1 or 20 PACs within "at least one year from the index encounter" it is counted at 1 (yes). I am not sure that this is reasonable. The PACs are also not weighted by severity.
- The composite performance measure appears to be overly broad, including numerous elements having little or no demonstrable connection with heart failure or the healthcare provided for patients with heart failure. Even if I were a lumper, instead of a splitter, I would not feel comfortable with this measure.

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications
2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about

the quality of care when implemented.

- The measure assesses the rate of patients with 1 or more PAC(s) during index episodes. This new risk adjusted outcomes measure is specified for use at the clinician group and team levels of analyses.
- The measure exclusively uses electronically available administrative claims data to calculation the measure score, and for this measure, better care equals lower scores.
- The developer describes non-patient-related PACs as controllable by provider processes without further analysis as other influencers that may contribute to PAC rates beyond the patient and provider (e.g., payer, access, suppliers, etc.).
- Patient- and claims- based exclusions are provided to promote the availability and consistency of claims data capture, including payer enrollment requirements, cost outliers of < 1% or > 99%, and claims not relevant to HF.
- Developers provide administrative claims data for HF & PAC (HF- & Patient Safety Failure-related) triggers, describe a complete 12-month episode time window. A calculation algorithm is provided.
- ICD-9 & ICD-10 codes are provided, though ICD-10 descriptions & an ICD-9 to ICD-10 crosswalk methodology are not provided.
- A conceptual risk model and statistical method using logistic regression model for determining the probability of a patient incurring a PAC are provided. After adjusting for patient-related factors, the developers state the remaining PAC variance is due to factors potentially controlled by the provider during and after hospitalization. "Predicted" coefficients from the risk adjustment models are summed to give predicted probabilities of PAC occurrence.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is this measure specified to pertain only to providers with at least 10 episodes (per the reliability testing described below)?*
- *Is it likely this measure can be consistently implemented?*
- *Is additional evidence required to determine whether group/practice/team level of analysis is appropriate?*

2a2. Reliability Testing Testing attachment

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.

- The developer tested reliability at the performance measure score, and used a beta-binomial model and a signal-to-noise analysis, which is appropriate for this type of measure, to differentiate the true difference between measured entities (the signal) to random measurement error (the noise). A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in between hospital performance. A value of 0.7 is often regarded as a minimum acceptable reliability value.
- The measure is specified for HF patients ≥ 18 years, though the testing sample includes patients 18 through 64 years
- Providers with < 10 HF episodes were excluded from reliability testing, though the measure is specified for patient without episode restrictions. A sample 6,025 HF episodes from 81 providers were included after less than 10 HF episodes were excluded. The mean age of 53.4 (18-64 years) and 38% female in the testing analysis exclusively using administrative claims data.
- The developer states, "Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that

minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset will apply to another." The developer also states that very high sample sizes are to achieve any meaningful and reliable comparisons.

- A patient may have more than one condition-specific concurrent episode with claims applied to both episodes. If an inpatient claim corresponds to a procedure index episode and to a condition index episode, the claim would be assigned to the procedure index episode, rather than the condition index episode (e.g., for a claim that corresponds to both index episodes of HF & CABG, the claim would be assigned to CABG).
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- Reliability results are provided in the table below, as well as in great detail in the accompanied spreadsheet with median (IQR) results demonstrating reliability of 0.61(0.52,0.75) for ≥ 10 providers, increasing with the number of providers, demonstrating the measure is able to demonstrate differences in performance. For reliability analysis, providers were restricted to the minimum of 10 HF episodes, though all episodes were included in the risk model.

Reliability Scores	Minimum # Episodes Per Provider		
	>=10	>=25	>=50
# of Providers (%)	81 (100)	27 (33)	13 (16)
Median (IQR)	0.61 (0.52,0.75)	0.80 (0.75,0.85)	0.85 (0.83,0.87)
Range	0.43-0.94	0.69-0.94	0.80-0.94

The table provides a summary reliability scores minimum sample size thresholds. Complete results are provided in the workbook entitled, NQF_HF_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab.

Questions for the Committee:

- *Reliability testing was conducted only for those providers with at least 10 episodes. Can differences in performance be identified for providers with fewer than 10 episodes? Should the measure be specified to include only those providers with at least 10 episodes? Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- Because this is an outcome measure, the rationale that is presented for subcriterion 1a does not necessarily have to address all of the variables used to calculate the measure.
- The measure uses a statistical risk model with 170 risk factors and episode-specific subtypes/severity markers including age, gender and clinical comorbidities, on at least 10 claims to determine "stable" covariates and assess comorbidity or procedure impact on the PAC. All covariates must be present prior to an episode trigger. No formal covariate analysis was conducted to select risk factors beyond the present on a minimum of 10 claims threshold. The developer describes the heterogeneity of the provided data sets as crucial to ensure measure validity, and the accuracy and completeness of the data sets.
- The developers did not provide disparities data, an exploration of a conceptual relation to SDS, or SDS factors in the risk model.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

○ *Are these variables available and generally accessible for the measured patient population?*

2b2. Validity testing

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.

- The developer conducted systematic assessment of face validity for the performance measure score for validity testing in numerous ways, including the use of monthly multi-specialty clinical working groups, and other tests of face validity, along with focus groups, face validity comparisons of the measure to other national accountability measures, as well as additional literature for the measure & PAC development process.
- No empiric results are provided for the face validity tests described above.

Questions for the Committee:

○ *Is the test sample adequate to generalize for widespread implementation?*

○ *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

○ *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer describes patient- (demographic, enrollment or low/high claims cost) and claims-based (due to missing or non-relevant data) exclusions for the measure. They further state nearly half of the original population of HF patients was removed from the denominator with applied exclusions.
- A significant number of patients were eliminated from the measure due to exclusion criteria, including 6,025 of 25,284 (23.8%) HF episodes (in 3,258,706 unique beneficiaries) and 81 of 2110 (3.8%) providers.

Questions for the Committee:

○ *Are high cost outliers (> 99%) exclusions an opportunity to identify PACs?*

○ *Does the high number of exclusions restrict the measure use?*

○ *Are the exclusions consistent with the evidence?*

○ *Are any patients or patient groups inappropriately excluded from the measure?*

○ *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- The risk model (detailed in the accompanied spreadsheet) includes 170 factors and episode-specific subtypes including age, gender, 12-month enrollment marker, co-morbidities, and episode severity markers.
- No SDS factors beyond age and gender was included in the risk-adjustment approach. Beyond noting that race was not available for analysis, no description of the of the conceptual relationships between patient sociodemographic factors, patient clinical factors, quality of care, and the outcomes (PAC rates) was provided., nor do they discuss the availability of SDS variables, beyond stating that "race" as an SDS variable was not available for analysis. The developer briefly discus general psychosocial and socioecononmic barriers that found in decreased HF care processes.
- Logistic regression was used to model the probability of at least one PAC during an episode. The reasoning for no additional modeling performed is described.
- The performance of the model was determined with a split sample method by estimating the model coefficients using a development dataset (80% of the sample) and applying those coefficients to a validation dataset (20% of the sample). C-statistics for the development and validation samples were 0.807 and 0.754, respectively. C-statistics measures the extent of a statistical model to discriminate between a patient with and without PAC, with an ability to predict if a PAC is or is not present about 69% to 74% of the time. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk

factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong.

- Both Hosmer-Lemeshow Goodness-of-Fit statistics and risk-decide plots were provided to indicate model fit. Results from the Hosmer-Lemeshow test suggest that the fit is not good; however, this test is sensitive to the number of groupings and sample sizes. Results from the risk decile plot indicate that the predicted PAC rates are similar to the observed PAC rates across all deciles of risk. The developer states the model demonstrates sufficient predictive power.

Questions for the Committee:

- *Is the Committee aware of conceptual relationship(s) between additional patient-level SDS factors, patient clinical factors, quality of care, and PACs (other than gender and age)? If so, what data might be available to allow an empirical analysis of these relationships?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*

2b5. Meaningful difference:

- The developer presents PAC rates across providers and also providers adjusting for differences in patient severity in a ratio of observed to expected attributable episodes to PACS accounting for patient severity, and calculates estimates from the risk model, for risk-standardized PAC rates for the provider.

Summary of Unadjusted and Adjusted Performance Scores Across Providers:

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25
Unadjusted		
Median (IQR)	41% (31%, 57%)	36% (23%, 42%)
Range	9%-80%	10%-79%
Adjusted (RSPR)*		
Median (IQR)	40% (32%, 46%)	37% (30%, 44%)
Range	14%-68%	14%-50%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_HF_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- As there is only one data source used for measure calculation (administrative claims), comparability of data sources or methods is not applicable.

Question for the Committee:

- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*

2b7. Missing Data

- No formal analysis of missing data is provided. As the measure assesses the rate of patients with PACs, rather

than the rate of PACs per index episode, the total number of PACs is not included in the PAC rate.
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- The developers state the under-coding of claims is unavoidable in the current DRG payment structure which could lead to under capture or missing PACs.

2d. Empirical Analysis to Support Composite Construction
- As an "any or none" composite, the individual complications are considered measurable components of the composite. Frequency and distribution statistics are provided in the PAC Overview and PAC Drill Down tabs in the measure spreadsheet, which detail PAC types and subtypes. The identification of individual PACs are not provided (e.g., sepsis, unattended falls, DVT).
- PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.
- The developer states that no formal analysis was performed on missing data. For details, see 2b7 above.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***
- Numerator -patients who triggered an episode of heart failure followed for AT LEAST ONE YEAR (vague) and had one or more PACs during the episode time window (not really defined)
- There are so many codes used for triggering and for PACs it would be nearly impossible to say that all the appropriate ones are included and inappropriate ones excluded.
- There are very complex decision tree and risk adjustment calculations made and based on claims data.
- The data elements are clearly defined individually, with appropriate codes included.
- The logic is clear, though I disagree with the inclusion of numerous elements.
- The beta-binomial method data for signal-to-noise ratio showed a wide range of scores across all providers. Among those with at least 10 episodes, the mean was 0.61, and scores for many were low. Scores among providers with 25 or more episodes were consistently good (mean = 0.80) and continued to improve as provider sample size increased to >50 (mean = 0.85). The developers assert that this demonstrates that for providers with a minimum number of episodes, the measure sufficiently differentiates performance.

2a2.: ***Committee's Comments on Reliability-Testing:***
- When testing for reliability, the developers included only providers with 10 or more episodes with PACs. Out of more than 2000 providers, the analysis only included 81. The reliability was still only 0.67 (average) for providers with >=10 episodes it was 0.61. There were only 27 providers with more than 25 episodes (Reliability 0.80).
- The good news is that not very many providers had a large number of episode that involved PACs. Just how well it differentiates between the >10 and more than 25 is hard to say.
- Differences in quality could be expected to be even larger in providers with <10 episodes, and should be measurable.
- The quality data is bound to be more valid for providers of >=10 episodes, so the measure should perhaps be specified for such providers.

2b1.: ***Committee's Comments on Validity-Specifications***:
- For many of the PACs, the evidence is only their own claims data that demonstrates at best an association between the PAC and heart failure.
- The specifications appear to be reasonably consistent with the evidence, though the validity as a composite measure is still troubling.

2b2.: ***Committee's Comments on Validity-Testing:***
- There are 170 risk factors used to adjust the risk any particular patient would have for a PAC.
- There is no disparities data but clearly disparities such as availability of care can be involved.
- The calculations for this measure are very complex and related to administrative claims data. The coding for the

claims is readily available to insurance carriers but the program for calculating the measure performance is not readily available.

- I am not convinced that the score for this measure is an indicator of quality. For example, electrolyte disturbance is a PAC but a conscientious provider would at least be looking for this problem by ordering blood work and then treating any abnormality -- which they would code for the visit and therefore be penalized for it.
- Content validity of the definitions of potentially avoidable complications (PACs) was established by working monthly with clinicians who are experts in their respective fields and specific to the episodes for which PACs are being measured. The clinical experts focused on whether or not a potentially avoidable complication can be deemed as such for a specific episode of care, and they helped to define and review all of the diagnosis and procedure codes for each PAC.
- Face validity was established in several ways. The developers presented results of claims data analyses that reveal the frequency and costs of PACs to physicians in several different healthcare systems involved in their pilot site implementations, as well as to medical directors from the employer coalitions and the health plans that provided the dataset to run the analyses. Some of these implementations include the Pennsylvania Employee Benefits Trust Fund and local provider groups and hospital, Horizon Blue Cross Blue Shield of NJ and many physicians and health systems.
- In addition, they performed dozens of analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. They state that "To-date, we have never experienced either wholesale or partial rejection of the results of analyses showing rates of PACs, which demonstrates the level of acceptability – face validity – of the measures from the payer, policymaker, employer and payer communities."
- The developers assert that measures of potentially avoidable complications have face-validity with consumers. In a series of focus groups, information presented about price and quality of certain providers was seen to influence the decisions of consumers. When the PAC measure was used, respondents selected the providers with the lowest PAC rates with a high level of confidence in choice, and used it as a surrogate for a strong quality signal.
- The test sample is adequate to generalize for widespread implementation.
- The results demonstrate sufficient internal validity for conclusions about quality to be made, though the external validity is still questionable.
- I do not agree that the score from this measure is a valid indicator of healthcare quality, because it incorporates too many PACs that cannot be influenced by improving healthcare quality in the real world.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Half or more of the denominator being removed based on exclusions seems like an awful lot.
- I do think that outliers are reasonable to exclude.
- The goodness of fit test revealed that there was not goodness of fit.
- I do not agree with the exclusion of high-cost outliers. They are especially likely to have PACs.
- The high number of exclusions definitely limits the usefulness of the measure.
- I am not clear whether the exclusions are consistent with the evidence or on why there were so many exclusions.
- No formal analysis was done on the impact of exclusions on performance scores.
- Risk adjustment for race and socioeconomic status probably would be appropriate, but the developers note that race was unavailable and they do not mention SES.
- The risk adjustment model variables are adequately described for measure implementation.
- No formal analysis of missing data was provided. Only missing gender was identified, with no other specific data types missing.
- No empirical data was provided to support the inclusion of individual elements into the composite measure, nor to demonstrate their added value. Although the aggregation is supported by the developer's rationale, I disagree with their rationale and consider it to be overly inclusive."

2d.: ***Committee's Comments on Composite Performance Measure:***

- I do not think that an all or none composite is a good measure of quality in this situation. Clearly some PACs have more impact than others. Serious PACs should be addressed before getting to the level of "any".
- This is no empirical analysis of the composite measure's performance beyond construct and face validity demonstrations.

Criterion 3. Feasibility

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

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than the person obtaining the original information (e.g., DRG, ICD-9 codes on claims).
- The developer provides an excel spreadsheet attachment including diagnoses, visits, hospitalizations, post-acute facility stays, procedures, laboratory tests and procedures/surgeries, for HF & PAC triggers, and describe the time window for measuring PACs as 12 months following a HF episode triggers, as well as a decision tree for measure calculation and implementation.
- The measure specifications, metadata and calculation algorithms are available free of charge on the developer's website. Limited analytics are planned at no cost to the end user.
- This is not an eMeasure.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- The data elements are available to claims analysts but not to other groups. The computer program with all of the risk adjustments is specific to this data set and may not apply to others.
- The developer does offer their calculation algorithm free of charge
- The required data elements are routinely generated and used during care delivery, and are available in electronic form.
- The data collection strategy is ready to be put into operational use.

Criterion 4: Usability and Use

4. Usability and 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.

- This is a newly developed claims measure is currently used in accountability programs for payers, states, and planned uses for public reporting, professional certification or recognition programs, and external quality improvement for benchmarking purposes.
- The developer states that PAC measures provide a foundation for the relationship between healthcare quality and cost and assist in the exploration of practice reengineering and alternative payment models, act as indicators of potential harm, and is spurring the development of private-based analytics software for further outcomes exploration.
- No public improvement rates are available due to recent implementation and variation in PAC definitions have also modified, though the provided PROMETHEUS data suggest wide variation in performance and improvement opportunities. Payer and Provider improvement use perspectives are outlined.
- The developer found no noted unintended consequences, though they state the measure is intended for transparency and QI activities only. They also state the under-coding of claims is unavoidable in the current DRG payment structure could be an unintended consequences of the measure, and payers calculating the measures even with inadequate

sample sizes and using the results to penalize providers could lead to invalid provider comparisons.

- If the measure was theoretically to be used for accountability purposes to "ding" the measured entity as defined in the level of analysis, further exploration of PAC antecedents and the measured entity is warranted, especially with small group practices and very small PAC rates. In the original testing sample of 2110 providers, when providers with fewer than 10 HF episodes were eliminated from analysis due to less reliability estimates with small numbers, 81 remained for analysis.

Questions for the Committee *:*

- *Is the measure publicly reported?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Should PAC measures also include the clinician: group in the analysis or include population-level only entities?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- The measure does not appear to be publically reported although it is in use and public reporting is "planned"
- The measure is publicly reported.
- The performance result could possibly be used to further the goal of high-quality, efficient healthcare.
- The PAC measures should include the clinician: group in the analysis and include population-level only entities.
- The potential benefits of the measure probably outweigh any potential unintended consequences (none of which were anticipated/specified), although the magnitude of potential benefit is quite unclear.

Criterion 5: Related and Competing Measures

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

0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, AHRQ) (endorsed)

- -CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html Summarize any harmonization efforts, i.e., responses from the developers regarding harmonization.
- Briefly summarize next steps according to protocol

Pre-meeting public and member comments

•

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2747

Measure Title: Proportion of Patients Hospitalized with Heart Failure (HF) that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)

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: 6/30/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☒ Health outcome: Potentially Avoidable Complications

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

The combination of the aging of the population and improved survival after acute myocardial infarction (AMI) has created a rapid growth in the number of patients currently living with chronic heart failure (CHF), with a concomitant increase in the number of hospitalizations for decompensated heart failure (McCoullough 2002). Despite advances in medical therapy, admission rates following heart failure hospitalization remain high (Stevenson 2011) (Cubbon 2011). Discharge from a heart failure hospitalization is followed by a readmission within 30 days in approximately 24% of cases (Desai 2012) with more than 50% patients readmitted to hospital within 6 months of discharge (Ross 2010). This is despite well-established guidelines like the, "Get With The Guidelines®- Heart Failure" the American Heart Association's collaborative quality improvement program of evidence-based care of patients hospitalized with heart failure (AHA 2013).

Causes of readmissions may be various. These may include local practice patterns, for example, hospitals with higher overall admissions tend to have higher readmission rates after HF hospitalization (Epstein 2011). Readmissions are also influenced by psychosocial and socioeconomic barriers that limit compliance with medications, life style changes, self-monitoring and appropriate follow up (Fonarow 2008). There is also a tendency for higher readmissions in centers with resource limitations such as lower nurse staffing levels and limited cardiac capabilities (Joynt 2011).

The need for consistently high quality, efficient care for heart failure is urgent. To improve accountability in the delivery of medical care, AHRQ has developed a list of patient safety indicators (PSIs) to identify potential harms to patients and a list of ambulatory care sensitive conditions (ACSCs) to identify admissions that could have been potentially avoided with good outpatient care (AHRQ 2008). Additionally, the Centers for Medicare and Medicaid Services (CMS) have taken a "Six Sigma" approach and defined Hospital Acquired Conditions (HACs) and "never events" that should almost never occur and are applying financial penalties when these events do occur (CMS 2012).

The Potentially avoidable complications (PAC) measure goes beyond the AHRQ PSIs, PQIs and the CMS HACs and creates a single comprehensive measure that measures all-cause harms for a patient with the index condition. Potentially avoidable complications (PACs) are the unwarranted health outcomes that this measure addresses (deBrantes 2010). Lack of patient education on self care techniques, diet and weight management; poor discharge instructions, poor care coordination, and poor arrangements of patient follow-up lead to unnecessary ER visits and hospitalizations due to acute exacerbations of heart failure and other complications (Bonow 2012). All these adverse events are aggregated together as a single comprehensive measure to study the overall rate of PACs in the HF population.

Adult patient diagnosed with Heart Failure

↓

Physician practices fail to educate patients / Physician practices have poor access

↓

Patient visits ER / gets hospitalized (Ambulatory care sensitive hospitalization event)

↓

Patient discharged with management advise / remains in hospital for treatment of PAC

Well-managed patients with HF should rarely incur a potentially avoidable complication such as an emergency room visit, and hospitalizations related to HF should be avoidable by proactive care and good management.

The enclosed workbook entitled NQF_HF_all_codes_risk_adjustment 06.30.15.xls lists the types of PACs and their frequency as calculated in a large regional database (see tab PAC overview). 42.5% of patients with HF had a PAC, with about 28% of PACs directly related to HF itself, such as respiratory insufficiency, acute CHF, pulmonary edema or fluid and electrolyte disorders (see tab PAC Drill Down Graph). Although the preventable hospitalizations in this dataset for the HF population were low, at only 5% of all HF episodes; approximately 26.5% of patients with HF had PACs related to patient centered care failures such as poor control of diabetes, urinary tract infections, and acute gastritis, many of them being managed in an outpatient setting in physician offices. As a result, about 41% of episodes had a PAC indicator on professional claims.

While PACs may not be eliminated completely, identifying the magnitude of PACs and knowledge of the cause for the most frequent or the most expensive PACs could place an emphasis in reducing them and as a consequence improving patient outcomes. The ability to clearly identify the type and frequency of each PAC creates a highly actionable measure for all providers that are managing or co-managing the patient; as well as for the health plan with whom the patient is a member (de Brantes 2009).

References:

1) McCoullough, PA, et al. “Findings from the Resource Utilization Among Congestive Heart Failure (REACH) study.” *Journal American College Cardiology* 39.1 (2002):60–69.

2)Stevenson LW, Pande R. “Witness to progress.” *Circulation* 4 (2011): 390–392.

3)Cubbon RM, et al. “Changing characteristics and mode of death associated with chronic heart failure caused by left ventricular systolic dysfunction: a study across therapeutic eras.” *Circulation* 4 (2011): 396–403. Web.

5) Desai, Akshay S., MD, MPH, Lynne W. Stevenson, MD “Rehospitalization for Heart Failure: Predict or Prevent?” *Circulation* 126 (2012): 501-506

6) Ross, JS, et al. “Recent national trends in readmission rates after heart failure hospitalization.” *Circulation* 3 (2010): 97–103. Web.

7) American Heart Association, and American Stroke Association. "Heart Failure Fact Sheet." *Get With the Guidelines Heart Failure* (2013): 1-4. Oct. 2013. Web.

8) Epstein AM, AK Jha, and EJ Orav. "The relationship between hospital admission rates and rehospitalizations." *N Engl J Med* 365 (2011): 2287–2295. Web.

9) Fonarow, GC, et al. "Factors identified as precipitating hospital admissions for heart failure and clinical outcomes: findings from OPTIMIZE-HF." *Arch Int Med* 168 (2008): 847-854. Web.

10) Joynt KE and AK Jha. "Who has higher readmission rates for heart failure, and why? Implications for efforts to improve care using financial incentives." *Circulation Cardiovascular Quality Outcomes* 4 (2011) 53–59. Web.

11) Department of Health and Human Services, Agency for Healthcare Research and Quality. 2008. ''AHRQ Quality Indicators. Prevention Quality Indicators: Technical Specifications, Version 3.2''

12) Centers for Medicare and Medicaid Services. "Hospital Acquired Conditions." *CMS.gov*. Centers for Medicare and Medicaid Services, 2012. Web.

13) de Brantes, Francois, Amita Rastogi, and Michael Painter. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." *Health Services Research 2nd ser.* 45.6 (2010): 1854-871. Web.

14) Bonow RO, et al. "ACCF/AHA/AMA-PCPI 2011 performance measures for adults with heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures and the American Medical Association-Physician Consortium for Performance Improvement." *Circulation* 125.19 (2012):2382-401.

15) de Brantes, François M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. "Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model." *NEJM* (2009) 361:1033 (Perspective)

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

Rationale: Better processes of care create an atmosphere of proactive management, consistency in care and standardized care patterns (Shekelle 2013) (Fenter 2006). Patient education and adopting safe practices significantly reduces occurrence of potentially avoidable complications (PACs) in all settings (Klein 2011) (Wachter 2013) (Berwick 2006) (Kovner 2011) (Farley 2013). It is known that by holding providers accountable for occurrence and costs of PACs, an built-in warranty is created around care of the index condition (de Brantes 2009).

While CHF has been noted as the most common indication for hospitalization for adults 65 and older, a prospective randomized trial showed a 56.2% reduction in the number of readmissions for heart failure due to intensive nurse-directed education, care coordination, and follow up (Rich 1995). This study also showed that the reduction of hospital admissions led to a savings on $460 per patient. Moreover, improved hospital and post-discharge care, including pre-discharge planning, home-based follow-up, and patient education, have all demonstrated decrease in heart failure related readmission rates suggesting that healthcare services / care processes influence outcomes in heart failure patients (Krumholz 2002).

AHRQ performed a meta-analysis of 53 published randomized control trials and reported on 47 studies. They found that home-visiting programs and heart failure clinic interventions, both of which are multicomponent complex interventions, reduced all-cause readmissions. However no single component intervention reduced all cause readmission. They also showed that interventions that focused on reducing readmissions did not impact mortality rates adversely (AHRQ 2014). Therefore if interventions are chosen according to the available body of evidence it is possible to reduce PAC's.

According to the Joint Commission (JC) heart failure performance measure, at discharge patients with HF must receive comprehensive discharge instructions that address activity level, diet, discharge medications, follow-up appointment, weight monitoring, and actions to take in case of worsening symptoms (Joint Commission 2010). These JC measures are publicly reported by hospitals. In 2011, the ACC/AHA/AMA (American Medical Association) Performance Consortium added a documented post-discharge appointment to the list of recommended HF performance measures (AHRQ 2014) (Bonow 2012).

Guideline directed medical therapy (GDMT) is a comprehensive combination of lifestyle modifications and medications and is tailored to care for a spectrum of HF patients ranging from outpatient care to hospitalized patients with a view to improving outcomes. GDMT if followed optimally should reduce hospitalizations in the former and reduce readmission rates in the latter. Various factors influence readmission rates and identifying these issues could give a window of opportunity to correct the trend. These may include local practice patterns, for eg. Hospitals with higher overall admissions tend to have higher readmission rates after HF hospitalization (Epstein 2011). Readmissions are also influenced by

psychosocial and socioeconomic barriers that limit compliance with medications, life style changes, self-monitoring and appropriate follow up (Fonarow 2008). There is also a tendency for higher readmissions in centers with resource limitations such as lower nurse staffing levels and limited cardiac capabilities (Joynt 2011).

Studies have demonstrated where care coordination exists, ambulatory care-sensitive hospitalizations decreased by 30% (Bodenheimer 2008). However, if patients do get hospitalized, discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013) (Mittler 2013). Another study from the Boston Medical Center, demonstrated that although one in five hospitalizations are complicated by post-discharge adverse events, development of a strong discharge services program for patients admitted for medical conditions reduced hospital utilization within 30 days of discharge (Jack 2009). In addition, while in the hospital, safe practices reduce the burden of healthcare associated complications (Ranji 2007). Some of these are listed below:

1. Frequent hand-washing reduce hospital acquired infections (WHO 2007)
2. Carefully implemented protocols lead to reduced line sepsis (Pronovost 2010)
3. Discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013)
4. DVT prophylaxis in patients on bed rest avoids pulmonary embolism (Shekelle 2013)
5. Frequent change in position of HF patients in the CCU avoids pressure sores (Shekelle 2013)

PAC measures in the setting of heart failure look at all-cause harms, such as the ones highlighted above, arising from poor management of a patient with heart failure.

References:

1) Shekelle, Paul G., et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

2) Fenter TC, et al. “The cost of treating the 10 most prevalent diseases in men 50 years of age or older.” *Am J Manag Care* 12.4 (2006): S90-S98.

3) Klein S. and D. McCarthy, OSF HealthCare: “Promoting Patient Safety Through Education and Staff Engagement.” *The Commonwealth Fund* 1475.7 (March 2011).

4) Wachter, Robert M., Peter Pronovost, and Paul Shekelle. "Strategies to Improve Patient Safety: The Evidence Base Matures." *Annals of Internal Medicine* 158.5_Part_1 (2013): 350-52. Web.

5) Berwick, D.M., D. R. Calkins, C. J. McCannon et al., “The 100,000 Lives Campaign: Setting a Goal and a Deadline for Improving Health Care Quality,” *Journal of the American Medical Association*,295.3 (2006):324–27

6) Kovner, Anthony R., James Knickman, and Steven Jonas. "Chapter 11." *Jonas & Kovner's Health Care Delivery in the United States*. 10th ed. New York: Springer Pub., 2011. Print.

7) Farley, TG, et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

8) de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? *Health Affairs (Millwood)* 28 (2009): w678-w687.

9) Rich MW, Beckham V, Wittenberg C et al. "A multidisciplinary intervention to prevent the readmission of elderly patients with congestive heart failure." *N Eng J Med* 333 (1995): 1190-5.

10) Krumholz HM, Amatruda J, Smith GL, et al. "Randomized trial of an education and support intervention to prevent readmission of patients with heart failure." *J Am Coll Cardiol* 39.1 (2002): 83-89.

11) *Welcome to the Effective Health Care Program*. Rockville, MD: U.S. Dept. of Health and Human Services, Agency for Healthcare Research and Quality, 2010. *Effective Health Care Program*. AHRQ, May 2014. Web.

12) The Joint Commission, Aide Joint Commission International, and WHO. "Improved Hand Hygiene to Prevent Health Care-Associated Infections." *World Health Organization*. *WHO Collaborating Centre for Patient Safety Solutions* (May 2007). Web.

13) Bonow RO, et al. "ACCF/AHA/AMA-PCPI 2011 performance measures for adults with heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures and the American Medical Association-Physician Consortium for Performance Improvement." *Circulation* 125.19 (2012):2382-401.

14) Epstein AM, AK Jha AK, EJ Orav. "The relationship between hospital admission rates and rehospitalizations." *N Engl J Med* 365 (2011): 2287–2295. Web.

15) Fonarow, GC, et al. "Factors identified as precipitating hospital admissions for heart failure and clinical outcomes: findings from OPTIMIZE-HF." *Arch Int Med 168* (2008): 847-854. Web.

16) Joynt KE, AK Jha. "Who has higher readmission rates for heart failure, and why? Implications for efforts to improve care using financial incentives." *Circulation: Cardiovascular Quality Outcomes* 4 (2011) 53–59. Web.

17) Bodenheimer, T. "Coordinating Care — a Perilous Journey through the Health Care System," *New England Journal of Medicine* 358.10 (2008): 1064-1071.

18) Weaver, Sallie J., Lisa H. Lubomksi, Renee F. Wilson, Elizabeth R. Pfoh, Kathryn A. Martinez, and Sydney M. Dy. "Promoting a Culture of Safety as a Patient Safety Strategy." *Annals of Internal Medicine* 158.5_Part_2 (2013): 369-75. Web.

19) Mittler, Jessica N., Jennifer L. O'hora, Jillian B. Harvey, Matthew J. Press, Kevin G. Volpp, and Dennis P. Scanlon. "Turning Readmission Reduction Policies into Results: Some Lessons from a Multistate Initiative to Reduce Readmissions." *Population Health Management 16*.4 (2013): 255-60. Web.

20) Jack BW, Chetty VK, Anthony D et al. "A Reengineered Hospital Discharge Program to decrease Rehospitalization: A randomized trial." *Ann Int Med* 150 (2009):178-187.

21) Ranji SR, Shetty K, Posley KA, et al. "Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol 6 Prevention of Healthcare-Associated Infections) Technical Review 9" *AHRQ* 04(07)-0051-6. January 2007.

22) Pronovost, P.J., G. A. Goeschel, E. Colantuoni et al., "Sustaining Reductions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study," BMJ 340:c309 (Feb. 4, 2010): Web.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
2747_HF_Evidence_Attachment_HCI3-635717853216006588.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 40% of its plan members with hypertension incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow- up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001) (NQF Quality Positioning System). Interestingly, in the

current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.

4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

6) BCBSNC: Blue Cross Blue Shield of North Carolina: https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5. http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/

11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.

14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology

Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The data included two years of administrative claims covering the period April 1, 2012 through December 17, 2014. There were a total 6,025 episodes of HF.

Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. After this exclusion 81 (out of 2110) providers remained. Performance scores of these providers are summarized in the following table:

Unadjusted PAC Rates:
- Median (IQR): 40.6% (30.8%, 57.1%)
- Range: 9.1% - 80%

Risk-Standardized PAC Rates (RSPR):
- Median (IQR): 39.9% (32.4%, 46.2%)
- Range: 14.5% - 67.9%

Please refer to the NQF_HF_all_codes_risk_adjustment 06.30.15.xls workbook under the "Provider Attribution" tab to see specific results for each provider.

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.
While CHF has been noted as the most common indication for hospitalization for adults 65 and older, a prospective randomized trial showed a 56.2% reduction in the number of readmissions for heart failure due to intensive nurse-directed education, care coordination, and follow up (Rich 1995). This study also showed that the reduction of hospital admissions led to a savings on $460 per patient. Moreover, improved hospital and post-discharge care, including pre-discharge planning, home-based follow-up, and patient education, have all demonstrated decrease in heart failure related readmission rates suggesting that healthcare services / care processes influence outcomes in heart failure patients (Krumholz 2002).

Many congestive heart failure (CHF) hospitalizations are considered potentially preventable and can be attributed to the failure of the outpatient health care system to properly manage and treat CHF (Will 2012). The American College of Cardiology Foundation (ACCF) and the American Heart Association (AHA) have jointly produced and updated guidelines for management of cardiovascular disease since 1980. The latest Heart Failure guideline released in 2013 states, "Adherence to the clinical practice guidelines herein reproduced should lead to improved patient outcomes" (Yancy 2013). These outcomes include reduction in PAC's due to HF including morbidity associated with progress of HF and a reduction in hospitalizations for HF.

The PAC measures go beyond simple readmission rates and look for all-cause harms in patients with heart failure. As our own analysis demonstrates, the readmission rates in heart failure patients are currently much lower (13.7%) than just a few years ago, as compared to published literature (Discharge from a heart failure hospitalization is followed by a readmission within 30 days in approximately 24% of cases (Desai 2012) with more than 50% patients readmitted to hospital within 6 months of discharge (Ross 2010)). Although this definitely suggests a positive trend in improvements over time, we do not have documented proof since our analysis is not on the same study population over time. On the other hand, we do notice that the overall PAC rates for the heart failure episode continue to be high (over 62%) and they are being reflected in high PAC counts in the professional claims (over 59%) suggesting that they have not completely been eliminated.

While PACs may not be completely eliminated, identifying their magnitude and understanding their causality, in particular for the most frequent or the most expensive, could lead to improving patient outcomes (de Brantes 2008) (de Brantes 2010).

References:
1) Rich MW, Beckham V, Wittenberg C et al. "A multidisciplinary intervention to prevent the readmission of elderly patients with

congestive heart failure." N Eng J Med 333 (1995): 1190-5.

2) Krumholz HM, Amatruda J, Smith GL, et al. "Randomized trial of an education and support intervention to prevent readmission of patients with heart failure." J Am Coll Cardiol 39.1 (2002): 83-89.

3) Will JC, AL Valderrama, and PW Yoon. "Preventable hospitalizations for congestive heart failure: establishing a baseline to monitor trends and disparities." Preventable Chronic Diseases 9.110260 (2012): Web.

4) Yancy, Clye W., MD, MSc, FACC, FAHA, Mariell Jessup, and Biykem Bozkurt. "2013 ACCF/AHA Guideline for the Management of Heart Failure." Circulation 128 (2013): 240-327. American Heart Association. Web.

5) Desai, Akshay S., MD, MPH, Lynne W. Stevenson, MD "Rehospitalization for Heart Failure: Predict or Prevent?" Circulation 126 (2012): 501-506

6) Ross, JS, et al. "Recent national trends in readmission rates after heart failure hospitalization." Circulation 3 (2010): 97–103. Web.

7) de Brantes F and Rastogi A. "Evidence-Informed Case Rates: Paying for Safer, More Reliable Care." The Commonwealth Fund 40 1146 (2008): 1-14.

8) de Brantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach". Health Services Research 45.6.2 (2010): 1854-1871.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Not applicable

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Heart failure (HF) is a chronic progressive resource intensive condition with a high economic and social burden. Approximately 5.1 million persons in the United States have clinically manifest HF, and the prevalence continues to rise (Go 2013). In the Medicare-eligible population, HF prevalence increased from 90 to 121 per 1000 beneficiaries from 1994 to 2003 (Curtis 2008). Heart Failure is one of the leading causes for hospitalization in Americans 65 and over and is the primary diagnosis in >1 million hospitalizations annually (Yancy 2013), accounting for a total Medicare expenditure exceeding $17 billion (Desai 2012). The total cost of HF care in the United States exceeds $30 billion annually, with over half of these costs spent on hospitalizations (Yancy 2013)(Go 2013). Patients hospitalized for HF are at high risk for readmissions, with a 25% rate of readmission within one month (Krumholz 2009).

In 2007, the Medicare Payment Advisory Commission called for hospital-specific public reporting of readmission rates, identifying HF as a priority condition. The Commission stated that readmissions for HF were common, costly, and often preventable (MedPac 2007). An estimated 12.5 percent of readmissions for HF were potentially preventable (AHRQ 2014). AHRQ has identified HF to be an ambulatory-care–sensitive conditions (ACSC) with a Prevention Quality Indicator (PQI) to track outcomes (Will 2012) (Ranji 2007).

The 2013 American Heart Association/American College of Cardiology (AHA/ACC) published guidelines for post discharge HF interventions (Yancy 2013). Despite advances in medical therapy across the spectrum of cardiovascular diseases, and the availability of guideline-directed medical therapy (GDMT) to represent optimal medical therapy as defined by ACCF/AHA guideline, hospitalizations due to poor management of HF continue to rise (Yancy 2013).

Many congestive heart failure (CHF) hospitalizations are considered potentially preventable and can be attributed to the failure of the outpatient health care system to properly manage and treat CHF (Will 2012). In 2012, CDC published a study analyzing data from National Hospital Discharge Survey between 1995 through 2009. There were 121,741 records with preventable hospitalizations for CHF among adults translating to a weighted number of 15,208,518 hospitalizations for adults in the United States during the 15-year study period, an average of 1,013,901 each year. Approximately 75% of preventable hospitalizations for CHF occurred in people aged 65 or older (Will 2012).

Therefore, there are many areas where improvement is possible in HF, making it a high priority aspect of health care. The PAC measures go beyond simple readmission rates and look for all-cause harms in patients with heart failure.

1c.4. Citations for data demonstrating high priority provided in 1a.3

1) Go AS, et al. “Heart disease and stroke statistics–2013 update: a report from the American Heart Association. Circulation.” American Heart Association 127 (2013): e6–245. Web.

2) Curtis, L.H., et al. “Incidence and prevalence of heart failure in elderly persons, 1994–2003.” Arch Intern Med 168 (2008):418-424. Web.

3) Yancy, Clye W., MD, MSc, FACC, FAHA, Mariell Jessup, and Biykem Bozkurt. "2013 ACCF/AHA Guideline for the Management of Heart Failure." Circulation 128 (2013): 240-327. American Heart Association. Web.

4) Desai, Akshay S., MD, MPH, Lynne W. Stevenson, MD “Rehospitalization for Heart Failure: Predict or Prevent?” Circulation 126 (2012): 501-506

5) Krumholz HM, et al. “Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission.” Circ Cardiovasc Qual Outcomes 2 (2009): 407–13. Web.

6) MedPac. Report to the Congress: Promoting Greater Efficiency in Medicare. Rep. Washington DC: Medicare Payment Advisory Commission, 2007. Web.

7) Welcome to the Effective Health Care Program. Rockville, MD: U.S. Dept. of Health and Human Services, Agency for Healthcare Research and Quality, 2010. Effective Health Care Program. AHRQ, May 2014. Web.

8) Will JC, AL Valderrama, and PW Yoon. “Preventable hospitalizations for congestive heart failure: establishing a baseline to monitor trends and disparities.” Preventable Chronic Diseases 9.110260 (2012): Web.

9) Ranji SR, Shetty K, Posley KA, et al. “Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. Technical Review 9 (Prepared by the Stanford University – UCSF Evidence-based Practice Center under Contract No. 290-02-0017).” AHRQ Publication No. 04(07)-0051-6. Rockville, MD: Agency for Health Research and Quality. January 2007.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*

Not applicable

1d. Composite Quality Construct and Rationale

1d.1. A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
    - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient); or
    - any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient).

1d.1. Please identify the composite measure construction: any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

1d.2. Describe the quality construct, including:

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

We classify PACs into two types: Type 1 PACs are directly related to the index condition and are often controlled by the servicing provider; Type 2 PACs, on the other hand result from patient safety failures and could be reduced by better systems and better processes in care. Both types of PACs could occur in any setting and so could be identified through any type of claims coming in the administrative dataset, including in-patient, out-patient, or professional claims. PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1.

The PAC measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs), Hospital Inpatient Quality Reporting measures, Avoidable Readmissions, AHRQ defined patient safety indicators (PSIs), NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates.

All defined PACs, irrespective of their type, or site of occurrence, are aggregated to create an overall comprehensive, composite measure. They all have equal weighting, since they are measured simply by the frequency of their occurrence.

1d.3. Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.

Each individual PAC, when measured in isolation, provides a very limited picture of the performance of the provider(s) who are managing or co-managing the care of the patient. However, looking at all the PACs that may occur individually or concurrently in a patient with a given episode provides a comprehensive picture of the care received by the patient for that particular condition or illness.

Additionally, the frequency of occurrence of individual PACs may be so low that it may require very high sample sizes from individual providers to achieve any meaningful and reliable comparisons. But aggregating all the PACs into a single quality metric creates meaningful scores that can be compared across providers even with relatively smaller sample sizes.

Additionally, a comprehensive measure is easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

1d.4. **Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**

In constructing the comprehensive composite PAC measure, each component PAC, as clinically defined by the subject matter experts, was given the same weight so that arbitrary weights may not bias the results. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence. As such, the patient is the ultimate unit of measurement and if the patient incurred any PAC during the episode, then that counts against the numerator.

Since the emphasis of the PAC measure was to simply identify the occurrence of PACs in any setting, aggregation of the PAC counts to create a comprehensive quality score with equal weights has been met with overall support from the clinical working groups as well as from the implementation sites.

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 subcriteria 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. **Cross Cutting Areas** *(check all the areas that apply):*

Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Healthcare Associated Infections, Safety : Medication Safety, Safety : Readmissions, Safety : Venous Thromboembolism

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.)*

http://www.hci3.org/ecr_descriptions/ecr_description.php?version=5.2.006&name=CHF&submit=Submit

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** NQF_HF_all_codes_risk_adjustment_06.30.15-635719668228888693.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.

Not applicable

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)*

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

Outcome: Number of patients who triggered an episode of heart failure (HF), are followed for at least one-year, and had one or more potentially avoidable complications (PACs) during the episode time window.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back*

to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)
The time window is the most recent 12 months of the episode, once a patient has triggered a HF episode.

S.6. **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, 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.
Patients that have triggered a HF episode, and are identified as having services for potentially avoidable complications (PACs), during the most recent 12 months of the episode. The enclosed excel workbook entitled NQF_HF_all_codes_risk_adjustment_06.30.15 gives the detailed codes for PACs in the tabs entitled PACs I-9 and PACs I-10. PACs are identified only based on diagnosis codes.

Services for PACs are identified as follows:
a. Any service (professional, outpatient facility, ancillary) that is relevant to HF and has a PAC code in any position on the claim
b. Any admission to an acute care facility, that is relevant to HF
c. Any admission to a post-acute care facility that is relevant to HF and has a PAC code in any position on the claim

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Adult patients aged 18 years and above who triggered an episode of heart failure (HF) and are followed for at least one-year.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Please refer to the enclosed excel workbook entitled
NQF_HF_all_codes_risk_adjustment 06.30.15

The target population is identified using the following criteria:
1. Using administrative claims database, patients with HF are identified using one of the following criteria:
a. Patients having an office visit with a trigger code of HF in any position, followed by a second confirmatory office visit (with a trigger code of HF in any position), at least 30 days apart.
b. Patients with a Principal Dx of a HF trigger code on an in-hospital stay claim.
The trigger codes for HF are provided in the tab called "Triggers I-9" or "Triggers I-10".
2. The patient should have continuous enrollment for the entire time window with no more than 30 days as an enrollment gap, with the entity providing the data (so we can ensure that the database has captured most of the claims for the patient during the episode time window).
3. The patient should have a complete episode time window in the claims data – so there are at least 12 months of claims in the database for the patient.
4. Patient should be at least 18 years of age

Once the episode is triggered all relevant claims are assigned to the episode. Relevant claims could be inpatient facility claims, outpatient facility claims, professional services, laboratory services, imaging services, ancillary claims, home health, durable medical equipment as well as pharmacy claims across the entire continuum of care centered around the patient's episode of care. Relevant claims are identified as those that have a diagnosis code that matches the codes in the typical Dx codes tabs (Typical Dx I-9 or Typical Dx I-10), or in the PAC Dx codes tab (PACs I-9 or PACs I-10) AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant admissions to acute and post-acute care facilities are also included in the denominator.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days

during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the HF measure if they are considered not relevant to HF care.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Denominator exclusions include exclusions of "patients" as well as "claims" not relevant to HF care. Please refer to the enclosed excel workbook entitled (NQF_HF_all_codes_risk_adjustment 06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
d. If the patient does not have at least 12 months of claims in the database (this helps eliminate incomplete episodes).
e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure if they meet one of the following criteria:
a. If none of the diagnosis codes on the claim are on the list of "triggers" or relevant diagnosis codes (either typical Dx or PAC Dx) for HF.
b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for HF.
c. If the HF trigger hospitalization also triggers a major surgical procedure such as coronary bypass procedure or angioplasty, suggesting that HF may be a comorbidity or an indication for the surgery.
d. The "principal" diagnosis on an inpatient stay claim during the episode time window triggers its own episode
e. The procedure code on a claim during the episode time window triggers its own episode

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
None

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Conceptual Model
Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, the remaining variance in PACs are due to factors that could be controlled by all providers that are managing or co-managing the patient, both during and after hospitalization.

Statistical Method:
Logistic Regression model to determine the probability of a patient incurring a PAC
Demographic variables, comorbid conditions, as well as clinical severity indicators are fed as independent risk factors into the model. Risk Factors are collected historically. Subtype information is collected from the index claim and any look-back period, if relevant. Subtypes are clinical severity indicators suggesting severity of the episode itself, for example, diagnosis of diastolic heart failure in a HF patient. For each patient the "predicted" coefficients from the risk adjustment models are summed to give the predicted probabilities of the occurrence of a PAC.

Risk Factors :(Please refer to the enclosed excel workbook entitled (NQF_HF_all_codes_risk_adjustment 06.30.15.xls). The risk factors along with their codes are listed in the tabs called “All Risk Factors I-9” and “All Risk Factors I-10” and also listed below:

AGE CONTINUOUS VARIABLE
GENDER FEMALE = 1 (MALE IS REFERENCE = 0)

Risk Factor #	Risk Factor Name
RF0101	Anoxic Brain Damage, persistent vegetative state
RF0102	Delirium, Meningitis, Encephalitis
RF0103	Previous Stroke, Paralysis
RF0104	Cerebral Palsy and Other Paralytic Syndromes
RF0105	Spinal Cord Disorders/Injuries
RF0106	Polyneuropathy
RF0107	Multiple Sclerosis
RF0108	Convulsions, Epilepsy
RF0109	Dementia
RF0110	Parkinson´s and Huntington´s Diseases
RF0111	Cerebrovascular Disease
RF0115	after care, rehabilitation
RF0201	visual loss, blindness, retinal tear, detachment
RF0301	ENT, Upper Respiratory Problems
RF0401	Respiratory Failure, O2, ventilator dependence
RF0402	Advanced COPD, Asthma
RF0403	Empyema, bronchiectasis, Pneumonias
RF0404	Aspiration Pneumonia, Laryngeal Problems
RF0406	TB, Pneumoconiosis, Aspergillosis
RF0407	Tobacco use, Lung disease due to External Fumes
RF0408	Other Lung Disease
RF0501	Previous Shock, Syncope, Vent Fibrillation
RF0503	Advanced CHF
RF0504	Cardiomyopathy, valve disorders
RF0505	Cardiac Arrhythmias, Heart Block
RF0506	Pacemaker, AICD
RF0507	Endocarditis, Other post surgical cardiac problems
RF0508	Other Cardiovascular Disease
RF0511	DVT, Pulm Embolism, Pulm Heart Disease
RF0512	Unstable Angina
RF0513	Hypotension, chronic, orthostatic
RF0514	Hyperlipidemia
RF0515	Intraaortic Balloon Pump
RF0516	ventricular assist device, ecmo, prolonged bypass
RF0517	Previous electrophysiology studies, cryoablation
RF0518	Recent AMI
RF0519	Previous PCI
RF0520	Previous CABG
RF0521	Previous Heart & Valve Surgery
RF0522	Previous aortic reconstruction
RF0523	Previos carotid endarterectomy
RF0524	Aortic and peripheral vascular disease
RF0525	Advanced Aortic and Vascular Disease
RF0601	GI Bleed
RF0602	Intestinal Obstruction/Perforation
RF0603	Acute Gastritis, Duodenitis

RF0604 Gastroduodenal Ulcer
RF0606 Intestinal Uro-genital Fistula
RF0607 Abdominal hernia w complications
RF0608 Vascular insufficiency of intestine
RF0609 Inflammatory Bowel Disease
RF0610 Irritable Bowel
RF0611 Diverticulitis, Meckel´s
RF0612 Digestive congenital anomalies
RF0613 Intestinal infection
RF0614 Esophageal Perforation, Hmg, Barretts, Compl Hiatal Hernia
RF0615 Abnormal weight loss
RF0616 Achalasia, Esophageal spasm, Stricture, Dysphagia
RF0617 GERD, Hiatal Hernia, Other Upper GI Disorders
RF0618 Previous Bariatric Surgery
RF0619 Hx of colon polyps, family Hx of colon cancer
RF0620 Enterostomy, GI devices, lap band
RF0701 Pancreatic Disease
RF0702 Perforation, fistula GB, bile duct, pancreas
RF0703 Gall stones, cholecystitis
RF0704 End-Stage Liver Disease
RF0705 Hepatitis, Cirrhosis, Other Hepatbiliary Disorders
RF0706 Recent Gall Bladder, Hepatobilary Surgery
RF0707 Acute Pancreatitis, pseudo cyst
RF0801 Bone/Joint/Muscle Infections/Necrosis
RF0802 Muscular Dystrophy
RF0803 Osteoporosis, ostetits deformans, pathological fracture
RF0804 Rheumatoid Arthritis and Inflammatory Connective Tissue Disease
RF0805 Gout and other crystal arthropathies
RF0806 Other arthropathies
RF0807 Osteoarthritis
RF0808 Joint Deformities
RF0809 Knee derangements
RF0810 Traumatic Dislocation Knee
RF0811 Dislocation Hip
RF0812 Synovitis, Ruture Tendon
RF0813 Status Knee Replacement
RF0814 Status Total Hip Replacement
RF0901 Decubitus Ulcer
RF0902 Skin and wound problems
RF1001 Diabetes, poor control
RF1002 Advanced diabetes
RF1003 diabetes
RF1101 Acute renal failure
RF1102 Dialysis Dependent
RF1103 Nephritis
RF1104 Chronic renal failure
RF1105 Urinary Tract Infections
RF1301 Endometriosis
RF1302 Fibroid uterus, benign tumors of female organs
RF1303 Pelvic Inflammatory disease
RF1304 Uterine prolapse, cystocele, vaginocele
RF1305 Female Harmonal Disorders
RF1306 Ovarian, Broad Ligament Disorders
RF1308 Other disorders of uterus, cervix

RF1309 Menopausal Disorders
RF1310 Menstrual Disorders
RF1401 Multiparity, multigravida
RF1402 Elderly Primi, other
RF1403 Poor obstetric history
RF1406 Cervical incompetence
RF1407 Abnormalities of uterus, female genital tract
RF1408 Hypertension, pre-eclampsia in Pregnancy
RF1409 Severe pre-eclampsia w HTN, Eclampsia
RF1410 Maternal, gestational diabetes, large for date
RF1411 Genital Herpes
RF1412 Infections of genitourinary tract, venereal disease in pregnancy
RF1413 Infectious Diseases in Mother
RF1414 Cardiovascular disease in Mother
RF1415 Mental Disorders in Mother
RF1416 Epilepsy in Mother
RF1417 Liver and biliary tract disorders in mother
RF1418 Kidney Disease in Mother
RF1419 Other Maternal conditions
RF1421 Cephalopelvic Disproportion due to maternal causes
RF1436 Peripartum Cardiomyopathy
RF1441 Previous Cesarean section
RF1450 Maternal Obesity, previous Bariatric Surgery
RF1454 Previous Rupture Uterus, Obstetrical Trauma
RF1458 Complicated Pregnancy Delivery
RF1460 Thrombophlebitis, DVT during Pregnancy
RF1461 Puerperal Sepsis, other major puerperal complications
RF1462 Obstetrical Embolism, Air, Amniotic Fluid, Pulm, Pyemic
RF1467 Tobacco Use in Mother
RF1601 Bleeding Disorders
RF1602 Severe Hematological Disorders
RF1603 Disorders of Immunity
RF1604 Nutritional and other Anemias
RF1605 Long-term use of anticoag, Aspirin
RF1701 Head and Neck Cancers
RF1702 Lung and Intrathoracic Cancers
RF1703 Neuroendocrine, Myeloproliferative Cancers
RF1704 Poorly differentiated, Secondary, Metastatic Cancers
RF1705 Other Tumors
RF1706 Acute Leukemia
RF1707 Cancer uterus, localized female organs
RF1708 Colorectal, Hepatobiliary and other GI cancers
RF1709 Breast, Prostate, Thyroid cancers
RF1710 Testicular Cancer and localized of male organs
RF1711 Cancer of Bladder and Urinary Tract
RF1712 Musculoskeletal Cancers
RF1801 Sepsis, MRSA, Opportunitistic infections
RF1901 Schizophrenia
RF1902 Major Depressive, Bipolar, and Paranoid Disorders
RF2001 Drug/Alcohol Psychosis
RF2002 Drug/Alcohol Dependence
RF2101 Drug Reactions, long term use of drugs
RF2102 Intra-abdominal injury
RF2201 Extensive Third-Degree Burns

RF2301 Major Organ Transplant Status
RF2302 Artificial Openings for Feeding or Elimination
RF2303 Complications of Medical & Surgical Care and Trauma
RF2304 severe morbid obesity
RF2305 morbid obesity
RF2306 obesity
RF2307 mild sleep apnea, hypoventilation
RF2308 moderate sleep apnea, hypoventilation
RF2309 obstructive sleep apnea
RF2310 Severe Protein-Calorie Malnutrition
RF2311 Mild-mod malnutrition
RF2401 Severe Head Injury
RF2402 Major Head Injury
RF2403 Vertebral Fractures without Spinal Cord Injury
RF2404 Falls, Fractures
RF2405 Amputation
RF2501 HIV/AIDS

Subtypes for HF
Diastolic Heart Failure
Cardiomyopathy
Hypertensive Heart Disease w Heart Failure
Hypertensive Heart Disease w Heart Failure & CKD
Heart Aneurysm and other Sequelae of AMI
Heart Valve Disorders
Previous heart valve replacement
Acute pericarditis
Chronic, adhesive, constrictive pericarditis
Other pericarditis
Myocarditis
Pulmonary heart disease
Other heart disease
Pacemaker, Defibrillator in place
Transplanted Heart
Protein Calorie Malnutrition
Tobacco Use

The prevalence of the risk factors in our reference dataset are listed in the enclosed workbook entitled NQF_HF_all_codes_risk_adjustment 06.30.15.xls – see tab "Risk Factor Prevalence". The output of the regression model are given in the same workbook in the tab 'Risk Model'.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Please see tab entitled Risk Model in the enclosed excel workbook entitled (NQF_HF_all_codes_risk_adjustment 06.30.15.xls).

S.16. **Type of score:**
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*
Please refer to the enclosed excel workbook entitled (NQF_HF_all_codes_risk_adjustment 06.30.15.xls).

Assembling the Denominator:

Using administrative claims database, patients with HF are identified using one of two criteria: 1) Patients having an office visit with a trigger code of HF in any position, followed by a second confirmatory office visit (with a trigger code of HF in any position), at least 30 days apart, 2) Patients a Principal Dx of a HF trigger code on an in-hospital stay claim. The trigger codes for HF are provided in the tab called “Triggers I-9” or “Triggers I-10”.

Patients are retained if they are 18 years of age or more, do not have a missing gender, have at least 12 month of claims in the database, have a maximum of 30-day enrollment gap for the entire episode time window, and have no outlier episode costs. All relevant professional, laboratory, imaging, ancillary and other claims that are incurred during the episode time window are included as part of the episode. Claims are considered relevant to HF care if they have one of the diagnosis codes, as listed on the tab entitled Triggers I-9, Triggers 1-10, PACs I-9, PACs I-10, Typical Dx I-9, or Typical Dx I-10 in any position on the claim AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant admissions to acute and post-acute care facilities are also included in the denominator. All relevant pharmacy claims carrying codes that match the ingredients listed in the Pharmacy tab of the enclosed workbook are also included as part of the episode. Services are pulled as part of the HF episode based on the diagnosis codes as defined above or if they have a service code that is marked as “sufficient” for that episode.

If a patient has more than one concurrent episode, and the claim is relevant to both episodes, the claim could get multi-assigned, except in the case of procedural episodes that get carved out with respect to the index stay. So if an inpatient stay claim carried a principal Dx code that matched the trigger diagnosis code for HF but they also had a procedure code for CABG (coronary artery bypass surgery), the stay claim would get uniquely assigned to CABG and not be counted with HF.

Once all the episodes are assembled, episodes that match the exclusion criteria, such as those with outlier costs, are flagged (those with total episode costs less than 1st percentile or greater than 99th percentile), and excluded from the final analysis.

Assembling the Numerator:

For every episode included in the denominator, services are flagged as having a PAC (potentially avoidable complication) based on the criteria listed below:

- Any service (professional, outpatient facility, ancillary) with a PAC code in any position on the claim
- Any admission to an acute care facility, that is relevant to HF

Relevant claims that do not have any PAC codes, and do not qualify as a PAC based on the criteria outlined above, are listed as typical claims. All included relevant pharmacy services are flagged as typical. Patients that have even a single PAC claim are counted as part of the numerator.

Calculating the measure:

Proportion of HF patients that have PACs is simply the ratio of patients with PACs within the HTN population and is called the PAC rate as shown in the equation below:

PAC rate = Patients with HF that have at least one PAC claim / Total number of HF patients

A flow chart demonstrating the series of steps and the counts of patients at each step is shown in tab entitled Decision Tree of the enclosed workbook called NQF_HF_all_codes_risk_adjustment 06.30.15.xls

Drill Down Calculations:

Further analysis from this construct helps create actionable reports.

For example as shown in the tab labeled PAC overview, not only do we have the PAC rate for a population, we can break them down by the PAC type – type 1 being directly related to HF and so actionable by the servicing physician, while type 2 PACs are related to patient safety failures and can be improved by process improvement. Additionally, analyzing potentially avoidable admissions helps focus strategies in reducing them.

Risk Adjustment:

Once we have the observed PAC rates, we risk-adjust them for patient factors such as patient demographics, comorbidities collected historically, and for severity of illness or procedure using subtypes collected from the index stay and / or look-back period. This helps adjust for factors outside the providers control and levels the playing field for provider performance comparisons.

Unit of Analysis:
The unit of analysis is the individual episode.

Dependent Variable:
The dependent variable is a dichotomous variable indicating whether an episode had one or more claims assigned as a PAC (=1) or not (=0).

Independent Variables:
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_HF_all_codes_risk_adjustment 06.30.15.xls

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., unstable anginadiastolic heart failure). Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_HF_all_codes_risk_adjustment 06.30.15.xls

As mentioned previously, to avoid creating perverse incentives all comorbidities and subtypes are identified prior to or at the very start of the episode. None are identified during the episode period.

Statistical Methods
We use logistic regression to model the probability of at least one PAC occurring during the episode. Only comorbidities and subtypes are included in the models as covariates if they are present in at least 10 episodes to prevent unstable coefficients. No further model building is conducted after the initial models are built. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but it does not make it a priority that all covariates in the model be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity of a episode condition, and lets each regression model determine for itself which of the factors are more significant for a specific episode. Non-significant covariates in episode models can not overly influence predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

When more than one line of business is included in the data, separate models are calculated for each sample (i.e., commercial, Medicaid etc.).

Provider Attribution and calculating PAC rates by provider:

Once episodes are constructed they are attributed to providers based on one of the various attribution rules. For HF, episodes are attributed to the primary care physician, internist, or cardiologist with the highest count of office visits.

Using the logistic regression technique described above, a model is developed that gives estimates for each risk factor and subtype for the patients in the population analyzed. These estimates are used to develop patient-level probabilities for the occurrence of PACs. The patient-level probability estimates are summed to construct aggregated measures (e.g., facility/provider-level). This method is similar to the methods employed by the Centers for Medicare and Medicaid Services (CMS) and endorsed by the National Quality Forum (NQF) to construct similar facility- and practice-level measures (i.e., mortality, readmissions, etc.):

For each provider, the number of actual observed occurrences of the outcome is summed across all attributed patients with that episode, to give the observed PAC rates for the provider.

Similarly adjusted probabilities from the risk adjustment models are summed across all attributed patients to give expected PACs for the provider.

The observed sum is then divided by the summed probabilities (O/E). This number yields whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). This calculation yields a practice-level unstandardized performance ratio.

To facilitate accurate comparisons of rates across units of analysis, this ratio is then standardized to the community rate using the indirect method. Specifically, the provider-level rate is multiplied by the expected community rate, calculated as the sum of adjusted probabilities for every individual in the sample across all providers in the analysis. This measure, known as the standardized rate, represents what the unit's risk-adjusted rate would be for the outcome of interest if its patient population was reflective of the of the overall community.

The formula for this calculation is as follows:

Adj Outcome_j={(SUM Observed_ij)/(SUM Prob_ij)} × {(SUM Prob_i) / (# of episodes)}
Where individual is attributed to unit of analysis j (e.g., practice, provider, etc.)

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset will apply to another.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If patient related data is missing, the case is deleted from both the numerator and the denominator

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The information is based on a two-year claims database from a large regional commercial insurer. The database has over 3.2 million covered lives and $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

The methodology can be used on any claims database with at least two years of data and a minimum of 150 patients with the index condition or hospitalization. Having pharmacy data adds to the richness of the risk-adjustment models.
The calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.
The methodology has been tested on databases of several health plans as well as on a few employer databases.

No data collection instrument was used.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual, Clinician : Team

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinician Office/Clinic, Other
If other: Across the care continuum

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2747_HF_Testing_Reliability_Validity_HCI3-635719668856800768.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2747
Measure Title: Proportion of Patients with Heart Failure (HF) that have a Potentially Avoidable Complication (during the episode time window)
Date of Submission: 06/30/15
Type of Measure:
Composite Construction:
☐Two or more individual performance measure scores combined into one score
☐ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)
☐ Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

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 <u>all</u> measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For <u>outcome and resource use</u> measures**, section **2b4** also must be completed.
- If specified for **<u>multiple data sources/sets of specificaitons</u>** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) 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 20 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.

<u>Note</u>: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.
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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. 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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The information is based on a two-year administrative claims database from a large regional commercial insurer. The database contains medical and pharmacy claims on over 3.2 million covered lives and more than $25.9 billion in "allowed amounts" for costs.

1.3. What are the dates of the data used in testing? April 1, 2012 – December 17, 2014

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.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Integrated Delivery System	☐ other:

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

There were a total of 2,110 providers in the data set. Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the reliability calculations. After this exclusion, there were 81 providers left.

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

After exclusions (see 2b.3.1 below), there were a total of 6,025 episodes of HF were included in the testing and analysis. Patients in these episodes were, on average, 53.4 years of age (range 18-64) and 38% were female. We did not have race information on these patients. All patients for this analysis had a trigger inpatient claim of HF as identified in our code tables.

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.

For the reliability analysis, we restricted the data to only providers with at least 10 attributed episodes. For risk adjustment, all episodes were used in the analysis, regardless of the provider to which they were attributed.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
None of the analyses included SDS variables.

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*)

We assessed the reliability of the measure to demonstrate that it sufficiently differentiates performance between providers using the beta-binomial method, which is applicable to measures of this type. Reliability is a measure that distinguishes the signal (the extent of performance variation between entities that is due to true differences in performance) from statistical noise. Our approach follows directly from the methods outlined in the technical report "The Reliability of Provider Profiling: A Tutorial" by J.L. Adams.

Reference:

Adams JL. The Reliability of Provider Profiling: A Tutorial. Rand Corporation. http://www.rand.org/pubs/technical_reports/TR653.html.

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*)

The table below provides a summary of the reliability score for different minimum sample size thresholds. For complete results, refer to the workbook entitled, NQF_HF_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab to see provider-specific results.

Reliability Scores	Minimum # Episodes Per Provider		
	>=10	>=25	>=50
# of Providers (%)	81 (100)	27 (33)	13 (16)
Median (IQR)	0.61 (0.52,0.75)	0.80 (0.75,0.85)	0.85 (0.83,0.87)
Range	0.43-0.94	0.69-0.94	0.80-0.94

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?*)

Reliability scores can vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or variation across patients within providers) whereas a reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (see Adams, 2009 cited above).

Although there was a wide range of scores across all providers with at least 10 episodes and scores for many were generally low, those among providers with 25 or more episodes were consistently good and continued to improve as provider sample size increased. This demonstrates that for providers with a minimum number of episodes the measure sufficiently differentiates performance.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Composite performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

☐ **Systematic assessment of content validity**

☐ **Validity testing for component measures** *(check all that apply)*

***Note**: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☐ **Endorsed (or submitted) as individual performance measures**

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

☐ **Empirical validity testing of the component measure score(s)**

☐ **Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. 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)*

Content validity was built into the development of the definitions of potentially avoidable complications (PACs). This involved working with clinicians who are experts in their respective fields and specific to the episodes for which PACs are being measured. In particular, the clinical experts focused on whether or not a potentially avoidable complication can be deemed as such for a specific episode of care, and help defined and review all of the diagnosis and procedure codes for each PAC. The enclosed link lists clinicians who have participated in the various Clinical Working Groups (http://www.hci3.org/content/clinical-working-group-contributors). Some of the clinical experts have also participated in monthly webinars that highlight the clinical aspects of these measures (http://www.hci3.org/content/using-ecrs-providers).

In addition, we illustrate that our measure has face validity in several ways.

Beyond the up front work performed by clinical experts, the validity of the measure has also been tested in various real world settings. For example, we have presented results of claims data analyses that reveal the frequency and costs of PACs to physicians in several different healthcare systems involved in our pilot site implementations, as well as to medical directors from the employer coalitions and the health plans that provided the dataset to run the analyses. Some of these implementations include the Pennsylvania Employee Benefits Trust Fund and local provider groups and hospital, Horizon Blue Cross Blue Shield of NJ and many physicians and health systems.

In addition, we have performed dozens of analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:

- Vermont Payment Reform Commission
- Maine Health Management Coalition
- WellPoint / Anthem CT
- NY State Medicaid
- CT Medicaid

- CO All-payer Claims Database, Center for Improving Value in Health Care

These analyses and their results have influenced, and continue to influence, the development of various public reporting, payment reform and delivery system reform efforts. To-date, we have never experienced either wholesale or partial rejection of the results of analyses showing rates of PACs, which demonstrates the level of acceptability – face validity – of the measures from the payer, policymaker, employer and payer communities.

As importantly, measures of potentially avoidable complications have face-validity with consumers. In a series of focus groups, Judy Hibbard and colleagues[1] examined the impact of presenting information about price and quality of certain providers in influencing the decisions of consumers. They tested the validity of PACs as a discriminator of quality, as well as other measures of quality, and used the dollar symbol to illustrate the level of price, much like is done for restaurant reviews. When the PAC measure was used, respondents selected the providers with the lowest PAC rates with a high level of confidence in choice, and used it as a surrogate for a strong quality signal. To the contrary, when more standard measures of quality were used, consumers tended to ignore them and use price as a surrogate for quality. As such, what the researchers found is that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Finally, our measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs)[2], Hospital Inpatient Quality Reporting measures [3], Avoidable Readmissions [4,5], AHRQ defined patient safety indicators (PSIs) [6], NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates [7].

References:

[1] Hibbard JH, Greene J, Sofaer S, Fiminger K, and Hirsh J. An Experiment shows that a well-designed report on Costs and Quality can help consumers choose High-Value Health Care. *Health Affairs* 2012; 31(3): 560-568. doi: 10.1377/hlthaff.2011.1168

[2] CMS defined Hospital Acquired Conditions: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

[3] CMS operated Hospital Inpatient Quality Reporting Program: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

[4] Jencks SF, Williams MV, and Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. N Engl J Med 2009 (Apr); 360 (14): 1418-1428. doi: 10.1056/NEJMsa0803563.

[5] Casalino LP, Pesko MF, Ryan AM et.al. Small Primary Care Physician Practices have low rates of Preventable Hospital Admissions. Health Affairs, 2014; 33(9): 1-9. doi: 10.1377/hlthaff.2014.0434.

[6] Agency of Healthcare and Quality defined Patient Safety indicators: http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx

[7] NQF endorsed measures: Quality Positioning System: http://bit.ly/1E5ZdP7

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Not applicable.

2b2.4. 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?*)

Given the significant clinical input that went into developing the measure, the widespread use and acceptance the measure has gained among a wide variety of individuals and organizations across the health system (public and private payers, clinicians, consultants, patients, etc.) [1-13], and the parallels between this measure and other measures that are in widespread use, this demonstrates that the measure has strong face validity.

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)

2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.

3. de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.

4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.

5. Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.

6. Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.

7. François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)

8. de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

9. Rastogi A, Mohr BA, Williams JO, Soobader MJ, de Brantes F. Prometheus Payment Model: Application to Hip and Knee Replacement Surgery. Clin Orthop Relat Res 2009; 467(10): 2587-2597.

10. de Brantes F and Rastogi A. Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. The Commonwealth Fund 40, publ. 2008; 1146:1-14.

11. de Brantes F, Gosfield A, Emery D, Rastogi A and G. D'Andrea, "Sustaining the Medical Home: How Prometheus Payment Can Revitalize Primary Care", Robert Wood Johnson Foundation Report, May 2009, http://www.rwjf.org/pr/product.jsp?id=42555, accessed October 2009.

12. de Brantes F, Camillus J. Evidence-informed case rates: a new health care payment model [Internet]. New York (NY): Commonwealth Fund; 2007 Apr [cited 2007 May 20]. Available from: http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=478278, Accessed Aug 1 2013.

13. Satin DJ, and Miles J. Performance Based Bundled Payments: Potential Benefits and Burdens. Available from: http://student.med.umn.edu/p4p-new/sites/default/files/MN%20Med%20Bundles%20Special%20Report%20-%20Satin.pdf, Accessed Aug 1 2013.

_____

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

Exclusions included exclusions of "patients" as well as "claims" not relevant to HF care. Please refer to the enclosed excel workbook entitled (NQF_HF_all_codes_risk_adjustment_06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
d. If the patient does not have at least 12 months of claims in the database (this helps eliminate incomplete episodes).
e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure if they meet one of the following criteria:
a. If none of the diagnosis codes on the claim are on the list of "triggers" or relevant diagnosis codes (either typical Dx or PAC Dx) for HF.
b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for HF.
c. If the HF trigger hospitalization also triggers a major surgical procedure such as coronary bypass procedure or angioplasty, suggesting that HF may be a comorbidity or an indication for the surgery.
d. The "principal" diagnosis on an inpatient stay claim during the episode time window triggers its own episode
e. The procedure code on a claim during the episode time window triggers its own episode

2b3.2. 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*)

We started with a total HF population of 25,284 episodes. After all the exclusions were applied, the remaining HF population included in the analysis consisted of 6,025 episodes. As mentioned above, no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers. As such, no formal exclusion testing was done.

2b3.3. 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*)

No formal analysis was done on the impact of exclusions on performance scores.

Descriptive Explanation:

Exclusions of patients were for the following reasons. Some are for comparative purposes and some for medical reasons.

(a) Comparative Purposes:
We excluded patients that did not have complete enrollment for the entire episode time window. This was done to ensure that the database had complete information on patients to be able to create the entire episode. Including patients with only a partial episode window could distort the measure by artificially reducing the actual count of patients with PACs.

(b) Medical Reasons:
Patients with outlier costs (less than 1st percentile value or greater than 99th percentile) were considered to be different from the general pool, and excluded from both the numerator and the denominator. This is another way to ensure that episodes are complete (because incomplete episodes may have very low costs), and do not bring in random noise into the analysis due to inappropriate codes or services (high outliers).

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1./S13 What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**
☐ **Statistical risk model with 170 potential risk factors** and episode specific subtypes
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.1.2. If an outcome or resource use 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.

2b4.2/S14. Identify the statistical risk model variables (Name the statistical method – e.g., logistic regression and list all the risk factor variables.
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_HF_all_codes_risk_adjustment_06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., Cardiomyopathy). Subtypes are specific to each unique episode and are included in the models only if they are present at the start of the episode. Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_HF_all_codes_risk_adjustment_06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Candidate comorbidities and subtypes were included in the models as covariates if they were present in at least 10 episodes to prevent unstable coefficients.

2b4.2.1/S15. Detailed risk model specifications including coefficients, equations, codes with descriptors, definitions(may be attached in an Excel or cvs file)
All Risk Factors with their coefficients are detailed in the enclosed workbook called NQF_HF_all_codes_risk_adjustment_06.30.15.xls – Please reference the tabs titled Risk Factor Prevalence and Risk Model.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Risk factors are comorbidity indicators collected from historical claims before the start of an episode. These are universally applied across all episodes. This list was selected based on input from clinical experts in clinical working groups. In addition, the Clinical Working Groups identified episode specific severity markers that were called episode subtypes and they help distinguish an episode as being more severe than another.

All risk factors and subtypes must be present prior to, or at the start of the episode and are identified using diagnosis codes in the patient's historical claims.

To be included in the risk adjustment models, any risk factor or subtype must be present in at least 10 episodes. Beyond this no further model building was conducted to add or remove risk factors or subtypes from the model after it was initially run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates cannot overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
As explained above, no formal analysis was conducted to select risk factors. In fact, all potential risk factors and subtypes with a count of at least 10 episodes were retained to serve as predictors. The goal was to achieve a more complete explanatory model rather than achieve parsimony.

Please reference the tabs titled Risk Model in the NQF_HF_all_codes_risk_adjustment_06.30.15.xls workbook to see the list of risk factors that met the selection criteria.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)
Not Applicable since our analysis did include SDS variables

2b4.5. 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*)

Model Development Approach

We used logistic regression to model the probability of at least one PAC occurring during the episode. The model included all covariates that were identified through the process above. No further model building was conducted after the initial model was run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates can not overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

For a more complete description of the risk adjustment approach, please see the document entitled, "PACs and Severity Adjustment Fact Sheet" that accompanies this submission.

Approach to Model Testing and Validation

To determine the validity and performance of the model, we used the split sample method to divide the patient sample randomly into: 1) the model building data set (80% of the sample) and 2) the test data set (20% of sample. The model was built using logistic regression on the first data set and then the coefficients from the development model were tested in the second dataset. Area under the curve (AUC) and the c-statistic were used to compare the predictive ability of the model in each of the data sets. Hosmer-Lemeshow Goodness-of-Fit tests and comparisons of observed to expected probabilities across risk deciles were further examined to assess the model's overall predictive accuracy.

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

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

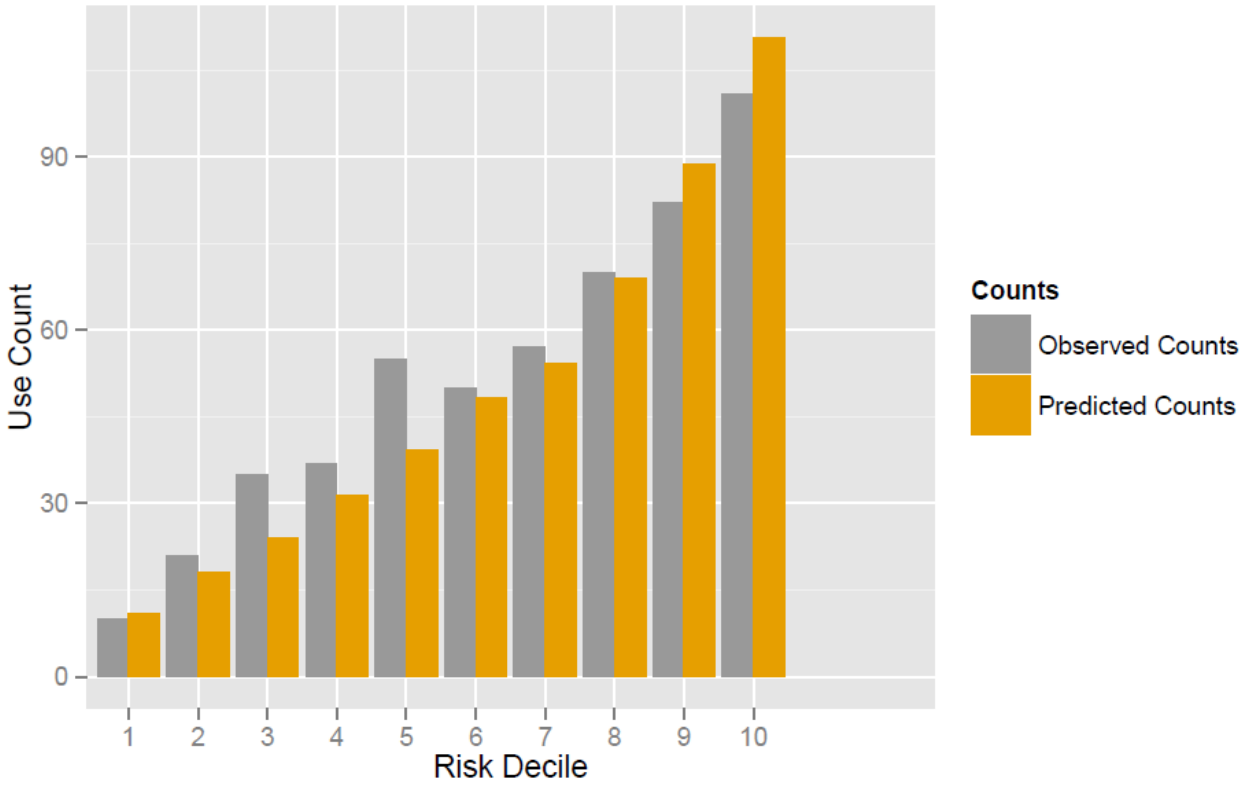
Sample	Accuracy (%)*	AUC
Test	73.5%	0.807
Validation	68.8%	0.754

*Episodes with predicted probabilities <50% were classified as having a predicted 0 (not having a PAC). Episodes with predicted probabilities >50% were classified as having a predicted 1 (having a PAC)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Sample	Chi Square	Degrees of Freedom	p-value
Test	31.6	8	<0.0001
Validation	30.6	8	<0.0001

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:



2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

The C statistic is a measure of the extent to which a statistical model is able to discriminate between a patient with and without an outcome. The c-statistic ranges from 0.5 to 1.0. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong [1]. The purpose of the model is to adjust for patient-related factors. The remaining unexplained differences in PAC rates are due to factors that could be controlled by all providers that are managing or co-managing the patient, during the entire episode time window.

The c-statistics of the testing and validation samples (0.807 and 0.754, respectively) indicate that the risk models have good discriminatory power. Indeed, the accuracy values show that the model correctly predicts whether an episode had or did not have a PAC 69% to 74% of the time, well above what would be expected if the predictions were made at random (i.e., 50%). Although the H-L test was significant for the testing sample, meaning that the model is not a good fit, this test is generally known to be sensitive to the number of groupings used and sample sizes. Nevertheless, the risk decile plot indicates, that, other than decile 5, the models predict PACs similarly to observed PACs across the risk deciles.

Overall, the results indicate the models have sufficient predictive power.

Reference:
[1] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Edition)*. New York, NY: John Wiley & Sons; 2000.

2b4.11. 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*)
Not applicable

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To directly compare PAC rates across providers while also appropriately accounting for differences in patient severity, we calculated a risk-standardized PAC rate for each provider. This method is similar to calculations used by others for reporting outcomes measures [1]. For each provider, the ratio of observed attributed episodes with PACs to the expected number of attributed episodes with PACs given the patient's risk factor and estimated from the risk-adjustment model was calculated. This number yielded whether the provider had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). We then multiplied this ratio by the overall expected PAC rate across all providers to obtain the risk-standardized PAC rate for the provider. This measure represents what a provider's PAC rate would be if its patient population was reflective of the overall population.

Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. Comparison of risk-adjusted PAC rates gives a measure of the provider's relative performance. Our analysis compared risk-standardized PAC rates across providers. We analyzed various descriptive statistics including the range in PAC rates, medians, interquartile range, etc.

References:

[1] See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

2b5.2. 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*)

Summary of Unadjusted and Adjusted Performance Scores Across Providers:

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25
Unadjusted		
Median (IQR)	41% (31%, 57%)	36% (23%, 42%)
Range	9%-80%	10%-79%
Adjusted (RSPR)*		
Median (IQR)	40% (32%, 46%)	37% (30%, 44%)
Range	14%-68%	14%-50%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_HF_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

2b5.3. 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?*)
Even after right-adjustment, the variation in risk-adjusted rates suggests there are meaningful differences in performance between providers in risk-standardized PAC rates for patients with an episode of HF.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note*: This criterion is directed 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 SDS 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.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

2b6.2. 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*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

If patient related data is missing, the entire patient is excluded from the numerator as well as the denominator.

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence.

According to our measure definition, in constructing the measure it is possible for a provider to have only one or some types of PACs and not others. Alternatively, the provider may have all PAC types occur for their patients. The measure only considers whether any PAC occurred regardless of the type, and all PAC types are weighted equally, therefore we believe, there is no potential for the absence of specific PAC types to bias performance scores for individual providers.

For these reasons, no formal analyses were done on missing data

2b7.2. 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*)
Not applicable

2b7.3. 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*)

Not applicable

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook entitled NQF_HF_all_codes_risk_adjustment_06.30.15.xls provides outputs from empirical analysis. The tab labeled "PAC overview" demonstrates percentage of episodes that had at least one PAC, and provides the breakdown of PACs: 1) by the type of PAC whether directly related to index condition or due to patient safety failures; 2) the setting of the PAC, whether seen in the in-patient setting, out-patient facility or during professional visits; and 3) preventable hospitalizations.

The "PAC Drill Down Graph" provides further detail on each component of the PAC and their frequency. As can be seen by the individual counts and the graph, while each individual PAC may have such small occurrences that no meaningful comparisons in provider performances could be made; together, they add value to provide a comprehensive picture that result in meaningful numbers. The aggregation of PACs to a comprehensive, composite measure, in itself provides the parsimony that is so desirable.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

All PACs, as clinically defined by the subject matter experts were used with equal weighting. Since the emphasis of the PAC measure is to identify the occurrence of PACs in any setting, a simple and straightforward approach was adopted.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc.*; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

No formal analysis was performed.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Additionally, if a patient had one or more PACs, it is simply counted as a 1, i.e., flagged as having a PAC. The measure only considers whether any PAC occurred regardless of the type, or the site, and all PAC types are weighted equally. Therefore, no formal analysis of individual components was performed.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

We chose not to weight the components of the measure.

Considerations were given to the fact that preventable hospitalizations may be given more weight, than PACs identified in a doctor's office. Similarly PACs in an in-patient setting may have more serious implications on a patient's ultimate outcome, than PACs occurring in an outpatient setting. Additionally, preventable hospitalizations as well as index hospitalizations, each with longer lengths of stay, may have serious PACs. But how do we weigh these effects? An alternative model was considered, where cost could be considered as a surrogate for the weights. Higher cost PACs could imply more serious PACs. However, differences in costs could be driven by many issues other than the PAC itself, such as unit price of the service, method of reimbursements, contracting arrangements etc.

Furthermore, in-patient facility billing does not allow for the distinction of PAC related costs from other costs within the stay. We would fail to capture PAC related costs within the stay and potentially underweight those. As a result, the decision was made to avoid weighting and keep the measure as a straightforward count.

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

Measuring all providers with the same yardstick will provide consistent results and reasonable comparisons over time. If the goal is to reduce PACs, then the PAC measure as was constructed with the help of various

experts in the field would provide reasonable comparisons. A word of caution however pertains to the sample size of the provider panel before making any reasonable conclusions.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved for high reliability in one dataset would apply to another.

2d3. Empirical analysis demonstrating that the approach for handling missing data minimizes bias (*i.e., achieves scores that are an accurate reflection of quality*)**.**

Note: *Applies to the overall composite measure; the focus is on missing data rather than exclusions, which are considered in 2b3.*

Please refer to section 2b7

2d3.1. What is the overall frequency of missing data and the distribution of missing data across providers?

2d3.2. Describe the method used to compare approaches for handling missing data (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2d3.3. What were the statistical results obtained from the analysis of missing data? (e.g., *results of sensitivity analysis of effect of various rules for missing data; if no empirical analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

2d3.4. What is your interpretation of the results in terms of demonstrating that the approach used for missing data minimizes bias? (i.*e., what do the results mean in terms of supporting the selected approach for missing data; if no empirical analysis, provide rationale for the selected approach for missing data*)

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in

electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
As part of our general implementation of these measures and related analyses, we have worked through dozens of different and sometimes very large datasets. From Medicare to Medicaid to regional and national commercial carriers, as well as individual employers, the principal lesson learned is the heterogeneity of the data sets and the significant variability in fill rate of critical data elements. As a result, we have created highly specific recommendations for which data elements are required to ensure measure validity, the accuracy of those data elements, and their completeness in the dataset. When claims datasets are organized in the way we specify in the measure analysis, and contain the coding information required, the analysis of the measure and its results are highly reliable.

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 calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Payment Program Blue Cross Blue Shield of North Carolina Blue Cross Blue Shield of New Jersey), Pennsylvania Employee Benefits Trust Fund https://www.bcbsnc.com/ http://www.horizonblue.com/ https://www.pebtf.org/ Quality Improvement (Internal to the specific organization) Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/

4a.1. **For each CURRENT use, checked above, provide:**

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

Measures associated to potentially avoidable complications (PACs) are in use today with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system (public and private payers, clinicians, consultants, all-payer claims database stewards, etc.) [1-8]. They are being used in various capacities in different pilot site implementations. To name a few:

•BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
•BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers

In addition, the PAC measures are incorporated by the following organizations in their bundled payment programs:

•BCBSSC – for CABG and PCI programs
•Horizon BCBSNJ– for CHF and CABG programs
•BCBSNC
•PEBTF in PA

http://www.ajmc.com/interviews/Lili-Brillstein-on-How-Bundled-Payments-Are-Tranforming-Healthcare

In these programs they look at PACs related to the measure for process improvement activities and for practice re-engineering.

We have created reports for rates of PACs for the following organizations:
-Vermont Payment Reform
-Maine Health Management Coalition
-WellPoint / Anthem CT
-NY State Medicaid
-CT Medicaid
-CO All-payer Claims Database, Center for Improving Value in Health Care

There are several companies that are leveraging these measures to create analytics and software for customers – these include HealthQx, Aver Informatics, McKesson, and TriZetto.

Below are some references that highlight our work with Potentially Avoidable Complications (PACs).

1.Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)

2.Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.
3.de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.
4.de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.
5.Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.
6.Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.
7.François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)
8.de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

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

4a.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.*)
Measures associated with PACs are currently in use as described in the prior section. In addition, we are working with several not-for-profit and for-profit organizations to provide them with the algorithms needed to calculate rates of potentially avoidable complications. Some of these organizations include:

Fair Health – based in NY and whose mission is to increase transparency of provider cost and quality,

CastLight – based in CA and serving large employers. We currently provide CastLight with Bridges To Excellence recognitions and will work with them to augment provider transparency by using PAC measures,

MA APCD (Massachusetts All Payers Claims Database) Council – we currently have an agreement in place with the MA APCD Council to produce PAC measures on hospitals and physicians and report back to the council with tests of reliability and validity of the measures. The purpose is to authorize the publication of these measures,
Maryland Health Care Cost Commission – we have a two year agreement to produce measures of cost and quality for public dissemination.

4b. 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.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

We do not have any public information to share about the improvements in rates of potentially avoidable complications, as the implementation of these measures is too recent to provide valid comparisons. Further, some of the definitions of PACs have changed since the measures were initially endorsed, making comparisons even more difficult and unreliable.

Nevertheless, the variation in performance scores presented in Section 1b.2 indicates that there are differences between providers in their risk-adjusted PAC rates (higher scores equal worse performance). This suggests that real opportunities exist to identify lower performing providers and reduce the overall occurrence of PACs.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Performance results provide summary PACs rates by provider, which can be used by payers and providers in a number of ways to improve the quality of care.

From the payer perspective, payers can use this information to 1) create a high-value provider networks, 2) work with high-value providers to share best practices, 3) incentivize low-value providers to improve, 4) modify their insurance design to activate consumers to select the right care from the right providers at the right time.

From the provider perspective, providers can 1) view services and activity for their patients longitudinally across the entire care continuum, such as frequency of readmissions and ED visits and drill down on patients with high PAC rates, 2) review actionable drill down reports to identify the most frequent PACs across all patients to create care pathways and process improvement plans to impact the most frequent PACs.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended consequences were reported, but there is the potential for:
1. Under-coding of PACs in the claim stream resulting in under-reporting the actual rate and/or providers gaming the measures
2. Payers calculating the measures even with inadequate sample sizes and using the results to penalize providers

The measure is designed for transparency efforts and to spur quality improvement. Detailed PAC reports can help providers identify areas of quality improvement. Even detailed reports of small samples of patients can be helpful for quality improvement purposes, but not for public reporting. To mitigate the potential for invalid provider comparisons, we specify in this submission the minimum sample size needed to ensure the reliability of a provider's score. Ultimately, there isn't any good way to prevent provider gaming of the measure by under-coding claims, however, under the current DRG payment methodology, many providers would be penalized by under-coding PACs since these codes often result in the assignment of more complicated DRGs.

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)**
0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the

30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
-0531 Patient Safety for Selected Indicators (Composite Measure, AHRQ) (endorsed)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Some of the measures listed in the prior section are, fully harmonized with the submitted measure, in particular, 0705, 0708, and 0709. Other measures such as 0337 and 0450 are in fact, subsets of our measure. However, there are some measures that are not harmonized, in particular the 30-day all-cause readmission measures and the Hospital wide all-cause readmission measure. While the submitted PAC measure include hospitalizations and readmissions that occur during the episode time window, the hospitalizations, by definition, have to be relevant to the underlying condition. For chronic conditions, most relevant hospitalizations within the entire episode time window are considered potentially avoidable. PACs include readmissions and are designed to enable accountability at the locus of provider control as well as some shared accountability between settings, centered around a patient, and for a specific medical episode of care. In that sense, they are consistent with the all-cause 30-day readmission rates, but represent a subset of those admissions. However, they do extend to the entire episode time window. As such, the PAC measures, as submitted, don't create added burden of reporting because the readmissions reported are simply a part of the broader 30-day all-cause readmission measures already endorsed by NQF. Because PAC measures are comprehensive, they include patient safety events as well as other adverse events, including hospitalizations and ED visits during the entire continuum of care. As a result, they are a comprehensive measure of avoidable complications for a specific medical episode. The data collection for all of the HCI3 measures is automated by a software package and is fully harmonized with all other PAC measures. A single download automates creation of all reports related to each of the PAC measures.

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.)
PAC measures are composite measures representing "all-cause harms". They look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence. PACs look at readmissions, emergency room visits, adverse events due to errors of omission or commission. They look at complications that are due to patient safety failures, and also

those directly related to the index condition. These are all a cause of significant waste and quality concerns. As such, the measure can provide clinicians with an overall and comprehensive view, in one measure, of all potentially avoidable complications for a patient and drive quality improvement efforts.

For clinicians and facilities increasingly engaged in value-based payment efforts and/or driving quality improvement for population health, the value of a PAC measure over a series of related, but more discrete measures, is that one can better determine if the sources of complications primarily stem from activities within the facility or outside the facility, and the specific nature of the complications that have a higher frequency of occurrence. While individual components of the PAC measure may have small frequencies and may be difficult to interpret with regards to provider performance or actionability, aggregating all the PACs into a comprehensive, composite measure provides the parsimony that is so desirable. For providers, it's far easier to construct a quality dashboard from a parsimonious set of measures, and that's what PAC measures offer.

Further, as a comprehensive outcome measure, PACs are also useful for public transparency of quality, as substantiated by the research from Judy Hibbard and colleagues previously cited in the "testing" section of this submission. As a comprehensive outcome measure, they are easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PACs_and_Severity_Adjustment_Fact_Sheet_HCI3.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.2 **Point of Contact:** Francois, de Brantes, Francois.debrantes@hci3.org, 203-270-2906-
Co.3 **Measure Developer if different from Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.4 **Point of Contact:** Amita, Rastogi, amita.rastogi@hci3.org, 213-934-9624-

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.
From 2006 onwards, and under the auspices of various funding organizations, HCI3 has convened and managed, or helped to convene and manage, Clinical Working Groups to inform the development and refinement of the measures. For example, in 2011, 2012 and 2013, HCI3 worked collaboratively with the American Board of Medical Specialties and the American Medical Association's Physicians Consortium for Performance Improvement, under a federal contract, to convene and get input from various clinical experts on definitions of episodes of care and their sequelae, including avoidable complications.

Some of the clinical experts that have contributed to our work include:
-Dr. John Allen, American Gastroenterology Association (AGA)
-Dr. Morton Arnsdorf, Cardiologist, University of Chicago, IL
-Dr. Peter Bach, Memorial Sloan Kettering Cancer Center (MSKCC)
-Dr. Peter Basch, Primary Care, Medstar Health, DC
-Dr. Justin Beckelman, Radiation Oncology, University of Pennsylvania, PA
-Dr. Debra Bingham, Executive Director, California Maternal Quality Care Collaborative (CMQCC) at Stanford University, CA
-Dr. John Birkmeyer, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Linda Bosserman, Wilshire Oncology Medical Group, CA

-Dr. Matthew Brengman, American Society of Metabolic and Bariatric Surgery (ASBMS)
-Dr. Joel Brill, American Gastroenterology Association (AGA)
-Dr. George Cautilli, Cautilli Orthopedic Surgical Specialists PC, Yardley, PA
-Dr. Ashwini Davison, Internist, Johns Hopkins Hospital, MD
-Dr. James Denneny, III, American Academy of Otolaryngology – Head and Neck Surgery (AAO-HNS)
-Dr. Chris Gallagher, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Robert Haralson, III, American Academy of Orthopedic Surgeons (AAOS)
-Ms. Dawn Holcombe, Executive Director, Connecticut Oncology Association, CT
-Dr. Colin Howden, American Gastroenterology Association (AGA)
-Dr. John Knightly, American Association of Neurological Surgeons (AANS)
-Dr. Larry Kosinski, American Gastroenterology Association (AGA)
-Dr. Nalini Krishnan, Obstetrics & Gynecology, MN
-Dr. Kelly Kyanko, Internist, NYU School of Medicine, NY
-Dr. Tara Lagu, Internist & Infectious Disease, Baystate Medical Center, MA
-Dr. Robert Lee, Society of Thoracic Surgeons (STS)
-Dr. Alex Little, Society of Thoracic Surgeons (STS)
-Dr. Michael London, Orthopedic Surgeon, OMNI Orthopedics, OH
-Dr. Elliott Main, Obstetrics & Gynecology, California Pacific Medical Center, CA
-Dr. Constantine Mantz, 21st Century Oncology, FL
-Dr. Joseph Messer, Cardiologist, Rush University Medical Center, IL
-Dr. David Metz, American Gastroenterology Association (AGA)
-Dr. Ronald Nahass, Infectious Disease Care, NJ
-Dr. Ajay Nehra, Urologist, Rush University Medical Center, IL
-Dr. Francis Nichols, Society of Thoracic Surgeons (STS)
-Dr. Patrick O'Connor, Primary Care, HealthPartners, MN
-Dr. Sara Perkel, National Comprehensive Cancer Network, PA
-Dr. David Peura, American Gastroenterology Association (AGA)
-Dr. John Ratliff, American Association of Neurological Surgeons (AANS)
-Dr. Steven Schutzer, Connecticut Joint Replacement Institute, CT
-Dr. Leif Solberg, Primary Care, HealthPartners, MN
-Dr. Scott Sporer, Midwest Orthopedics at Rush, Chicago IL
-Dr. Bonnie Weiner, Cardiologist, Worcester Medical Center, MA
-Dr. Jonathan Weiner, Bariatric Surgery codes, Prof of Health Policy and Management, Johns Hopkins University, MD
-Dr. Janet Wright, Cardiologist, Northstate Cardiology Consultants, CA

Measure Developer/Steward Updates and Ongoing Maintenance

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

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

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

Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

Measure Information - Composite

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 subcriterion 1b).

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

Brief Measure Information

NQF #: 2748

De.2. **Measure Title:** Proportion of Patients with Hypertension (HTN) that have a Potentially Avoidable Complication (during the episode time window)

Co.1.1. **Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)

De.3. **Brief Description of Measure:** Percent of adult population aged 18 + years who triggered an episode of hypertension (HTN), are followed for at least one-year, and have one or more potentially avoidable complications (PACs). PACs may occur any time during the episode time window. Please reference attached document labeled NQF_HTN_all_codes_risk_adjustment_06.30.15.xls, in the tabs labeled PACs I-9 and PAC I-10 for a list of code definitions of PACs relevant to HTN.

We define PACs as one of two types:

(1) Type 1 PACs - PACs directly related to the index condition: Patients are considered to have a PAC, if they receive services during the episode time window for any of the complications directly related to HTN, such as for malignant hypertension, blurred vision, acute CHF etc.

(2) Type 2 PACs - PACs suggesting Patient Safety Failures: Patients are also considered to have a PAC, if they receive services during the episode time window for any of the complications related to patient safety failures such as for sepsis, infections, phlebitis, deep vein thrombosis, pressure sores etc..

All relevant admissions in a patient with HTN are considered potentially avoidable and flagged as PACs.

PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook labeled NQF_HTN_all_codes_risk_adjustment_06.30.15.xls serves as an example. The tab labeled PAC overview gives the percent of HTN episodes that have a PAC and the tab labeled "PAC drill down" gives the types of PACs and their frequencies in HTN episodes within this dataset.

The information is based on a two-year claims database from a large regional commercial insurer. The database had over 3.2 million covered lives and over $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

1b.1. **Developer Rationale:** Each individual PAC, when measured in isolation, provides a very limited picture of the performance of the provider(s) who are managing or co-managing the care of the patient. However, looking at all the PACs that may occur individually or concurrently in a patient with a given episode provides a comprehensive picture of the care received by the patient for that particular condition or illness.

Additionally, the frequency of occurrence of individual PACs may be so low that it may require very high sample sizes from individual providers to achieve any meaningful and reliable comparisons. But aggregating all the PACs into a single quality metric creates meaningful scores that can be compared across providers even with relatively smaller sample sizes.

Additionally, a comprehensive measure is easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

S.4. **Numerator Statement:** Outcome: Number of patients who triggered an episode of hypertension (HTN), are followed for at least

one-year, and had one or more potentially avoidable complications (PACs) during the episode time window.
S.7. **Denominator Statement:** Adult patients aged 18 years and above who triggered an episode of hypertension (HTN) and are followed for at least one-year.
S.10. **Denominator Exclusions:** Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the HTN measure if they are considered not relevant to HTN care.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

Is this a MAINTENANCE measure submission? ☐ **Yes** ☒ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: n/a **Most Recent Endorsement Date:** n/a

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for this *health outcomes* measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if health outcomes measures agree the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

- This new risk-adjusted (by age, gender and clinical co-morbidities) outcomes measure that assesses the proportion of adult patients with claims triggered Hypertension (HTN) with at least one Potentially Avoidable Complications (PAC) within 12 months of CAD triggered claims data. Based on NQF's criteria, this measure is considered an "any or none" composite measure that assesses if 1 or more PACs or "care defects" have occurred for the index episode. For this composite measure, the individual complications considered the measurable components.
- PACs are classified in two types: 1) related to HTN, and 2) related to Patient Safety Failures, combining the 2 types into a single PAC rate to calculate the proportion of patient with 1 or more PAC. PACs are considered unwarranted health outcomes that combine concepts from AHRQ PSIs, PQIs and the CMS HACs and episode-specific PACs into all-cause patient harms that is measured during an index condition for use at the practice, medical group, provider system or purchaser/payer levels to identify quality of care gaps between practices and hospitals.
- The developer links primary & secondary prevention care gaps, poor patient education, poor care coordination and poor follow-up increase unnecessary ER visits, hospitalizations, readmissions, and mortalities to increased PACs, and state that PACs for HTN patients should occur rarely in well-managed patients.
- The evidence for Patient Safety Failure PACs is described to be within the influence of the measured entity, and does not describe the influence of non-healthcare-related impacts on PAC rates. The progression of the episode condition, illness or disease is also not mentioned as a contributor to PAC rates in the evidence.
- In addition to linking processes of care to outcomes, the developer provides an extensive PAC literature review in

sections 1a.2. and 1a.2.1. for HTN, Patient Safety Failures & processes of care, as well as background information on the process for PAC development.

Questions for the Committee:

- *Does sufficient evidence exist connecting Patient Safety Failures to the HTN index episode?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*

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

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

- The developer provides HTN prevalence & impacts data, rationales and general information on PAC measure utility and applicable setting use.
- HTN PAC performance gap data are calculated from PROMETHEUS administrative claims data from April 1, 2012 through December 17, 2014, for providers with ≥ 10 attributable index episodes. The data includes 3,702 of 23,125 (16%) providers from 262,273 of 409,442 index episodes in 3,258,706 unique beneficiaries.

Unadjusted PAC Rates:		**Risk-Standardized PAC Rates (RSPR):**	
Median (IQR):	29.7% (20.0%, 42.9%)	Median (IQR):	31.1% (24.2%, 37.8%)
Range:	0% -100%	Range:	0% -173%

- Limited descriptive data on the patient, provider and payer are provided. The developer provides "Overview" and "Drill Down" PAC rates in the spreadsheet demonstrating gap, though PAC rates for individual complications are not provided.
- The developer cites 2013 CDC National Health Statistics data on the prevalence & treatment of HTN, remaining constant at ~ 30% over the last decade, with Mexican-Americans born outside the US, persons without health insurance had lower BP control, with Black adults have higher HTN rates and lower BP control to Caucasians.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - Somewhat weak and unconventional for traditional CV measures

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- Somewhat weak and unconventional for traditional CV measures

1b. ***Committee's Comments on Performance Gap:***

- Very difficult to understand the performance gap as it captures one very common condition and relation of that condition to numerous, likely quite unrelated "adverse" outcomes

1c. ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure assesses the rate of patients with 1 or more PAC(s) during index episodes. This new risk adjusted outcomes measure is specified for use at the clinician group and team levels of analyses.
- The measure exclusively uses electronically available administrative claims data to calculation the measure score, and for this measure, better care equals lower scores.
- The developer describes non-patient-related PACs as controllable by provider processes without further analysis as other influencers that may contribute to PAC rates beyond the patient and provider (e.g., payer, access, suppliers, etc.).
- Patient- and claims-based exclusions are provided to promote the availability and consistency of claims data capture, including payer enrollment requirements, cost outliers of < 1% or > 99%, and claims not relevant to HTN.
- Developers provide a robust data set of administrative claims codes for HTN & PAC (HTN- & Patient Safety Failure-related) triggers, describe a complete 12-month episode time window. A calculation algorithm is provided.
- ICD-9 & ICD-10 codes are provided, though ICD-10 descriptions & an ICD-9 to ICD-10 crosswalk methodology are not provided.
- A conceptual risk model and statistical method using logistic regression model for determining the probability of a patient incurring a PAC are provided. After adjusting for patient-related factors, the developers state the remaining PAC variance is due to factors potentially controlled by the provider during and after hospitalization. "Predicted" coefficients from the risk adjustment models are summed to give predicted probabilities of PAC occurrence.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is this measure specified to pertain only to providers with at least 10 episodes (per the reliability testing described below)?*
- *Is it likely this measure can be consistently implemented?*
- *Is additional evidence required to determine whether group/practice/team level of analysis is appropriate?*

2a2. Reliability Testing Testing attachment

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.

- The measure is specified for HTN patients ≥ 18 years, though the testing sample includes patients 18 through 64 years
- The developer tested reliability at the performance measure score, and used a beta-binomial model and a signal-to-noise analysis, which is appropriate for this type of measure, to differentiate the true difference between measured entities (the signal) to random measurement error (the noise). A value of 0 indicates that all variation is due to

measurement error and a value of 1 indicates that all variation is due to real differences in between hospital performance. A value of 0.7 is often regarded as a minimum acceptable reliability value.

- A sample of 262,273 providers was initially included in the data set, though providers with less than 10 HTN episodes were excluded, allowing for 3,702 (1.4%) remaining providers. There were 147,169 HTN episodes with a mean age of 53.0 (18-64 years) and 46% female in the testing analysis exclusively using administrative claims data.
- A patient may have more than one condition-specific concurrent episode with claims applied to both episodes. If an inpatient claim corresponds to a procedure index episode and to a condition index episode, the claim would be assigned to the procedure index episode, rather than the condition index episode (e.g., for a claim that corresponds to both index episodes of HF & CABG, the claim would be assigned to CABG).
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- Reliability results are provided in the table below, as well as in great detail in the accompanied spreadsheet with median (IQR) results demonstrating reliability of 0.79 (0.67,0.89) for ≥ 10 providers, increasing with the number of providers, demonstrating the measure is able to demonstrate differences in performance. For reliability analysis, providers were restricted to the minimum of 10 HTN episodes, though all episodes were included in the risk model.

Reliability Scores	Minimum # Episodes Per Provider		
	>=10	>=25	>=50
# of Providers (%)	3,702 (100)	2,011 (54)	1,039 (28)
Median (IQR)	0.79 (0.67, 0.89)	0.87 (0.81, 0.92)	0.92 (0.89, 0.95)
Range	0.49-1.00	0.71-1.00	0.83-1.00

The table provides a summary reliability scores minimum sample size thresholds. Complete results are provided in the workbook entitled, NQF_HTN_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab.

Questions for the Committee:

- *Reliability testing was conducted only for those providers with at least 10 episodes. Can differences in performance be identified for providers with fewer than 10 episodes? Should the measure be specified to include only those providers with at least 10 episodes? Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- Because this is an outcome measure, the rationale that is presented for subcriterion 1a does not necessarily have to address all of the variables used to calculate the measure.
- The measure uses a statistical risk model with 170 risk factors and episode-specific subtypes/severity markers including age, gender and clinical comorbidities, on at least 10 claims to determine "stable" covariates and assess comorbidity or procedure impact on the PAC. All covariates must be present prior to an episode trigger. No formal covariate analysis was conducted to select risk factors beyond the present on a minimum of 10 claims threshold. The developer describes the heterogeneity of the provided data sets as crucial to ensure measure validity, and the accuracy and completeness of the data sets.
- The developers provide disparities data relation to age, race, and insurance status, though further exploration of a conceptual relation to SDS, or SDS factors in the risk model is not provided.

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Are these variables available and generally accessible for the measured patient population?*
- *Does the Committee find a conceptual relation between the provided disparities data and potential SDS factors?*

2b2. Validity testing

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.

- The developer conducted systematic assessment of face validity for the performance measure score for validity testing in numerous ways, including the use of monthly multi-specialty clinical working groups, and other tests of face validity, along with focus groups, face validity comparisons of the measure to other national accountability measures, as well as additional literature for the measure & PAC development process.
- No empiric results are provided for the face validity tests described above.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer describes patient- (demographic, enrollment or low/high claims cost) and claims-based (due to missing or non-relevant data) exclusions for the measure. They further state nearly half of the original population of HTN patients was removed from the denominator with applied exclusions.
- A significant number of patients were eliminated from the measure due to exclusion criteria, including 147,169 of 409,442 (35.9%) HTN (48.6%) episodes (in 3,258,706 unique beneficiaries) and 3,702 of 262,273 (1.4%) providers.

Questions for the Committee:

- *Are high cost outliers (> 99%) exclusions an opportunity to identify PACs?*
- *Does the high number of exclusions restrict the measure use?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- The risk model (detailed in the accompanied spreadsheet) includes 170 factors and subtypes including age, gender, 12-month enrollment markers, co-morbidities, and episode severity markers.
- The developers provide disparities data relation to age, race, and insurance status, though further exploration of a conceptual relation to SDS, or SDS factors in the risk model is not provided.
- Logistic regression was used to model the probability of at least one PAC during an episode. The reasoning for no additional modeling performed is described.
- The performance of the model was determined with a split sample method by estimating the model coefficients using a development dataset (80% of the sample) and applying those coefficients to a validation dataset (20% of the sample). C-statistics for the development and validation samples with c-statistic results of 0.800 and 0.801, respectively. C-statistics measures the extent of a statistical model to discriminate between a patient with and without

PAC, with an ability to predict if a PAC is or is not present about 80% of the time. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong.

- Both Hosmer-Lemeshow Goodness-of-Fit statistics and risk-decide plots were provided to indicate model fit. Results from the Hosmer-Lemeshow test suggest that the fit is not good; however, this test is sensitive to the number of groupings and sample sizes. Results from the risk decile plot indicate that the predicted PAC rates are similar to the observed PAC rates across all deciles of risk. The developer states the model demonstrates sufficient predictive power.

Questions for the Committee:

- *Is the Committee aware of conceptual relationship(s) between additional patient-level SDS factors, patient clinical factors, quality of care, and PACs (other than gender and age)? If so, what data might be available to allow an empirical analysis of these relationships?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*

2b5. Meaningful differences:

- The developer presents PAC rates across providers and also providers adjusting for differences in patient severity in a ratio of observed to expected attributable episodes to PACS accounting for patient severity, and calculates estimates from the risk model, for risk-standardized PAC rates for the provider.

Summary of Unadjusted and Adjusted Performance Scores Across Providers:

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25
Unadjusted		
Median (IQR)	30% (20%, 43%)	29% (20%, 41%)
Range	0%-100%	3%-100%
Adjusted (RSPR)*		
Median (IQR)	31% (24%, 38%)	31% (25%, 36%)
Range	0%-173%	4%-112%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_HTN_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- As there is only one data source used for measure calculation (administrative claims), comparability of data sources or methods is not applicable.

Question for the Committee:

- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*

2b7. Missing Data

- No formal analysis of missing data is provided. As the measure assesses the rate of patients with PACs, rather than the rate of PACs per index episode, the total number of PACs is not included in the PAC rate.
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- The developers state the under-coding of claims is unavoidable in the current DRG payment structure which could lead to under capture or missing PACs.

2d. Empirical Analysis to Support Composite Construction

- As an "any or none" composite, the individual complications are considered measurable components of the composite. Frequency and distribution statistics are provided in the PAC Overview and PAC Drill Down tabs in the measure spreadsheet, which detail PAC types and subtypes. The identification of individual PACs are not provided (e.g., sepsis, unattended falls, DVT).
- PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.
- The developer states that no formal analysis was performed on missing data. For details, see 2b7 above.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- The elements are well defined
- The problem I have is that these are associations not causations
- The adverse outcomes are associated with HTN but unlikely to be caused by HTN

2a2.: ***Committee's Comments on Reliability-Testing:***

- Probably reliable if examined in very large populations

2b1.: ***Committee's Comments on Validity-Specifications***:

- Not valid, see above.

2b2.: ***Committee's Comments on Validity-Testing:***

- Not valid
- These are putative associations not causation

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Not Applicable

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than the person obtaining the original information (e.g., DRG, ICD-9 codes on claims).
- The developer provides an excel spreadsheet attachment including diagnoses, visits, hospitalizations, post-acute facility stays, procedures, laboratory tests and procedures/surgeries, for HTN & PAC triggers, and describe the time window for measuring PACs as 12 months following a HTN episode triggers, as well as a decision tree for measure calculation and implementation.
- The measure specifications, metadata and calculation algorithms are available free of charge on the developer's website. Limited analytics are planned at no cost to the end user.
- This is not an eMeasure.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- Not feasible.

Criterion 4: Usability and Use

4. Usability and 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.

- This is a newly developed claims measure is currently used in accountability programs for payers, states, and is planned for public reporting, professional certification or recognition programs, and external quality improvement for benchmarking purposes.
- The developer states that PAC measures provide a foundation for the relationship between healthcare quality and cost and assist in the exploration of practice reengineering and alternative payment models, act as indicators of potential harm, and is spurring the development of private-based analytics software for further outcomes exploration. No public improvement rates are available due to recent implementation and variation in PAC definitions have also modified, though the provided PROMETHEUS data suggest wide variation in performance and improvement opportunities. Payer and Provider improvement use perspectives are also outlined.
- The developer found no noted unintended consequences, though they state the measure is intended for transparency and QI activities only. They also state the under-coding of claims is unavoidable in the current DRG payment structure could be an unintended consequences of the measure, and payers calculating the measures even with inadequate sample sizes and using the results to penalize providers could lead to invalid provider comparisons.
- If the measure was theoretically to be used for accountability purposes to "ding" the measured entity as defined in the level of analysis, further exploration of PAC antecedents and the measured entity is warranted, especially with small group practices and very small PAC rates. In the original testing sample of 262,273 providers, when providers with fewer than 10 HTN episodes were eliminated from analysis due to less reliability estimates with small numbers, 3,702 remained for analysis.

Questions for the Committee:

- *Is the measure publicly reported?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Should PAC measures also include the clinician: group in the analysis or include population-level only entities?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- I suspect this is one of the ""big data"" measures that can demonstrate association between almost any chronic condition and adverse outcome
- Whether there is causation and scientific validity should be proven first prior to implementing this type of measures

Criterion 5: Related and Competing Measures

.1a. List of related or competing measures (selected from NQF-endorsed measures)

0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)

- -CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2748
Measure Title: Proportion of Patients with Hypertension (HTN) that have a Potentially Avoidable Complication (during the episode time window)
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: 6/30/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☒ Health outcome: Potentially Avoidable Complications

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Given the ever-increasing number of patients with one or more chronic illness, the need for consistently high quality, efficient chronic illness care is urgent. While there is a general understanding of the nature of care failures in chronically ill patients (e.g. ambulatory care sensitive hospitalizations) (Yuen 2004), there has been no attempt to measure the magnitude or the type of potentially avoidable complications, and the cost reductions that would ensue if a payment model encouraged care to be optimized at benchmarks achieved in studies.

There is enough evidence in the literature that highlights significant "gaps in care" in management of patients with chronic conditions (McGlynn 2003). Gaps in care, in turn lead to process failures that cause patients to incur unnecessary services and some harm (Jha 2013). For example, a report by the Agency of Health Care Research and Quality (AHRQ) highlighted the fact that in 2008, $4.4 million out of a total of 39 million (11 percent) hospital-stays that could have been prevented (Stranges 2008); and for Medicare beneficiaries one in five admissions were for a potentially preventable condition (Jiang 2006). To improve accountability in the delivery of medical care, AHRQ has developed a list of patient safety indicators (PSIs) to identify potential harms to patients and a list of ambulatory care sensitive conditions (ACSCs) to identify admissions that could have been potentially avoided with good outpatient care (AHRQ 2008). Additionally, the Centers for Medicare and Medicaid Services (CMS) have taken a "Six Sigma" approach and defined Hospital Acquired Conditions (HACs) and "never events" that should almost never occur and are applying financial penalties when these events do occur (CMS 2012).

The Potentially avoidable complications (PAC) measure goes beyond the AHRQ PSIs, PQIs and the CMS HACs and creates a single comprehensive measure that measures all-cause harms for a patient with the index condition. Potentially avoidable complications (PACs) are the unwarranted health outcomes that this measure addresses (de Brantes 2010). Lack of patient education on self care techniques, poor care coordination, and poor arrangements of patient follow-up could lead to unnecessary ER visits, hospitalizations and gaps in care leading to increased morbidity. All these adverse events are aggregated together as a single comprehensive measure to study the overall rate of PACs in the HTN population.

Adult patient diagnosed with Hypertension
↓
Physician practices fail to educate patients / Physician practices have poor access
↓
Patient visits ER / gets hospitalized (Ambulatory care sensitive hospitalization event)
↓
Patient discharged with management advise / remains in hospital for treatment of PAC

Well-managed patients with HTN should rarely incur a potentially avoidable complication such as an emergency room visit, and hospitalizations related to HTN should occur only in the rarest of circumstances.

The enclosed workbook entitled NQF_HTN_all_codes_risk_adjustment_06.30.15.xls lists the types of PACs and their frequency as calculated in a large regional database (see tab PAC overview). Over 31% of patients with HTN had a PAC, with about 13% of PACs directly related to HTN itself, such as malignant hypertension, syncope, or fluid and electrolyte disorders (see tab PAC Drill Down Graph). Although the preventable hospitalizations in the HTN population were low, at only 1.0% of all HTN episodes; approximately 24% of patients with HTN had PACs related to patient centered care failures such as poor control of diabetes, respiratory insufficiency and acute gastritis, many of them being managed in an outpatient setting in physician offices. As a result over 30% of episodes had a PAC indicator on their professional claims.

While PACs may not be eliminated completely, identifying the magnitude of PACs and knowledge of the cause for the most frequent or the most expensive PACs could place an emphasis in reducing them and as a consequence improving patient outcomes. The ability to clearly identify the type and frequency of each PAC, creates a highly actionable measure

for all providers that are managing or co-managing the patient; as well as for the health plan with whom the patient is a member (de Brantes 2009).

References:

1) Yuen, E. ‘‘Severity of Illness and Ambulatory Care-Sensitive Conditions.’’ *Medical Care Research and Review* 61.3 (2004): 376–91.

2) McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. ‘‘The Quality of Health Care Delivered to Adults in the United States.’’ *New England Journal of Medicine* 348.26 (2003): 2635–45.

3) Jha, A. K., I. Larizgoitia, C. Audera-Lopez, N. Prasopa-Plaizier, H. Waters, and D. W. Bates. "The Global Burden of Unsafe Medical Care: Analytic Modeling of Observational Studies." *BMJ Quality & Safety* 22.10 (2013): 809-15. Web.

4) Stranges, Elizabeth, MS, and Carol Stocks, RN, MHSA. “Potentially Preventable Hospitalizations for Acute and Chronic Conditions, 2008.” *Agency for Healthcare Research and Quality* 99 (2010). Web.

5) Jiang HJ, Russo CA, and Barrett ML. Nationwide Frequency and Costs of Potentially Preventable Hospitalization, 2006. AHRQ-HCUP Statistical Brief # 72. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb72.jsp.

6) Department of Health and Human Services, Agency for Healthcare Research and Quality. 2008. ‘‘AHRQ Quality Indicators. Prevention Quality Indicators: Technical Specifications, Version 3.2’’ Available at http://www.qualityindicators.ahrq.gov

7) Centers for Medicare and Medicaid Services. "Hospital Acquired Conditions." *CMS.gov*. Centers for Medicare and Medicaid Services, 2012. Web.

8) de Brantes, Francois, Amita Rastogi, and Michael Painter. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." *Health Services Research 2nd ser.* 45.6 (2010): 1854-871. Web.

9) de Brantes, Francois M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. “Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model.” *NEJM* (2009) 361:1033 (Perspective)

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

Rationale: Better processes of care create an atmosphere of proactive management, consistency in care and standardized care patterns (Shekelle 2013) (Fenter 2006). Patient education and adopting safe practices significantly reduces occurrence of potentially avoidable complications (PACs) in all settings (Klein 2011) (Wachter 2013) (Berwick 2006) (Kovner 2011) (Farley 2013). It is known that by holding providers accountable for occurrence and costs of PACs, an built-in warranty is created around care of the index condition (de Brantes 2009).

Specifically for Hypertension, Staessen et.al, showed that active treatment with antihypertensive medication reduced the risk for stroke by 42%, cardiovascular endpoints by 31% and sudden death by 26% (Staessen 1997). Beckett et.al., in a large randomized control trial comprising 3,845 patients showed a reduction in fatal and nonfatal stroke rates when appropriate antihypertensive treatment was instituted (Beckett 2008). In addition to pharmacological treatment, appropriate life style modifications also play an important role in reducing PAC’s

due to hypertension, particularly cardiovascular morbidity. The Life style Work Group published their guidelines in 2013, with recommendations on nutrition and physical activity to control hypertension and reduce cardiovascular risk (Eckel 2013).

Studies have demonstrated where care coordination exists, ambulatory care-sensitive hospitalizations decreased by 30% (Bodenheimer 2008). However, if patients do get hospitalized, discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013) (Mittler 2013). Another study from the Boston Medical Center, demonstrated that although one in five hospitalizations are complicated by post-discharge adverse events, development of a strong discharge services program for patients admitted for medical conditions reduced hospital utilization within 30 days of discharge (Jack 2009). In addition, while in the hospital, safe practices reduce the burden of healthcare associated complications (Ranji 2007). Some of these are listed below:

1. Frequent hand-washing reduce hospital acquired infections (WHO 2007)
2. Carefully implemented protocols lead to reduced line sepsis (Pronovost 2010)
3. Discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013)
4. DVT prophylaxis in patients on bed rest avoids pulmonary embolism (Shekelle 2013)
5. Frequent change in position of HTN patients in the CCU avoids pressure sores (Shekelle 2013)

PAC measures in the setting of hypertension look at all-cause harms, such as the ones highlighted above, arising from poor management of a patient with hypertension.

References:

1) Shekelle, Paul G., et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

2) Fenter TC, et al. "The cost of treating the 10 most prevalent diseases in men 50 years of age or older." *Am J Manag Care* 12.4 (2006): S90-S98.

3) S. Klein and D. McCarthy, OSF HealthCare: "Promoting Patient Safety Through Education and Staff Engagement." *The Commonwealth Fund* 1475.7 (March 2011).

4) Wachter, Robert M., Peter Pronovost, and Paul Shekelle. "Strategies to Improve Patient Safety: The Evidence Base Matures." *Annals of Internal Medicine* 158.5_Part_1 (2013): 350-52. Web.

5) Berwick, D.M. D. R. Calkins, C. J. McCannon et al., "The 100,000 Lives Campaign: Setting a Goal and a Deadline for Improving Health Care Quality," *Journal of the American Medical Association*,295.3 (2006):324–27

6) Kovner, Anthony R., James Knickman, and Steven Jonas. "Chapter 11." *Jonas & Kovner's Health Care Delivery in the United States*. 10th ed. New York: Springer Pub., 2011. Print.

7) Farley, TG, et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

8) de Brantes, F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? *Health Affairs (Millwood)* 28 (2009): w678-w687.

9) Staessen JA, Fagard R, Thijs L, et al; "The Systolic Hypertension in Europe (Syst-Eur) Trial Investigators. Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension." *Lancet*. 350.9080 (1997;350):757-764.

10) Beckett, NS, Peters R, Fletcher AE, et al; "HYVET Study Group. Treatment of hypertension in patients 80 years of age or older." *N Engl J Med* 358.18 (2008):1887-1898.

11) Eckel, RH, Jakicic JM, Ard JD, et al. "AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association task force on practice guidelines." *Circulation* 2013. Web.

12) Bodenheimer, T. "Coordinating Care — a Perilous Journey through the Health Care System," *New England Journal of Medicine* 358.10 (2008): 1064-1071.

13) Weaver, Sallie J., Lisa H. Lubomksi, Renee F. Wilson, Elizabeth R. Pfoh, Kathryn A. Martinez, and Sydney M. Dy. "Promoting a Culture of Safety as a Patient Safety Strategy." *Annals of Internal Medicine* 158.5_Part_2 (2013): 369-75. Web.

14) Mittler, Jessica N., Jennifer L. O'hora, Jillian B. Harvey, Matthew J. Press, Kevin G. Volpp, and Dennis P. Scanlon. "Turning Readmission Reduction Policies into Results: Some Lessons from a Multistate Initiative to Reduce Readmissions." *Population Health Management 16*.4 (2013): 255-60. Web.

15) Jack BW, Chetty VK, Anthony D et al. "A Reengineered Hospital Discharge Program to decrease Rehospitalization: A randomized trial." *Ann Int Med* 150 (2009):178-187.

16) Ranji SR, Shetty K, Posley KA, et al. "Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol 6 Prevention of Healthcare-Associated Infections) Technical Review 9" *AHRQ* 04(07)-0051-6. January 2007.

17) The Joint Commission, Aide Joint Commission International, and WHO. "Improved Hand Hygiene to Prevent Health Care-Associated Infections." World Health Organization. WHO Collaborating Centre for Patient Safety Solutions (May 2007). Web.

18) Pronovost, P.J., G. A. Goeschel, E. Colantuoni et al., "Sustaining Reduc¬tions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study," BMJ 340:c309 (Feb. 4, 2010): Web.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:
1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?
N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
2748_HTN_Evidence_Attachment_HCI3.docx

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., the benefits or improvements in quality envisioned by use of this measure)*
Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 40% of its plan members with hypertension incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow- up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.

4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

6) BCBSNC: Blue Cross Blue Shield of North Carolina: https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5. http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/

11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.

14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

The data included two years of administrative claims covering the period April 1, 2012 through December 17, 2014. There were a total 262,273 episodes of HTN.

Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. After this exclusion 3,702 (out of 23,125) providers remained. Performance scores of these providers are

summarized in the following table:

Unadjusted PAC Rates:

Median (IQR):	29.7% (20.0%, 42.9%)
Range:	0% -100%

Risk-Standardized PAC Rates (RSPR):

Median (IQR):	31.1% (24.2%, 37.8%)
Range:	0% -173%

Please refer to the NQF_HTN_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

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.

National Center for Chronic Disease Prevention and Health Promotion and the National Center for Health Statistics, CDC in 2013 published its statistics on the prevalence and treatment of hypertension in the US. According to the report, the prevalence of hypertension has remained consistent at about 30% in the last decade. Among adults with hypertension, Mexican-American persons born outside the United States, and persons without health insurance had lower rates of blood pressure control in 2005–2008. Black adults have higher rates of hypertension and lower rates of blood pressure control compared to Caucasians (Gillespie 2013). Another significant finding was that although the prevalence of hypertension was lowest among those aged 18–44 years (9.8%), the prevalence of blood pressure control was significantly lower among this group than their older counterparts. This is most likely because of lower rates of hypertension awareness and treatment with medication among younger adults. In order to address this gap, the United States Preventive Services Task Force (USPSTF) recommends blood pressure screening for all adults' aged =18 years (USPSTF 2007) .

References

1) Gillespie, Cathleen D., MS, and Kimberly A. Hurvitz, MHS. "Prevalence of Hypertension and Controlled Hypertension — United States, 2007–2010." Centers for Disease Control and Prevention. National Center for Health Statistics, 22 Nov. 2013. Web. <http://www.cdc.gov/mmwr/preview/mmwrhtml/su6203a24.htm>.

2) US Preventive Services Task Force. "Screening for High Blood Pressure: U.S. Preventive Services Task Force Reaffirmation Recommendation Statement." Annals of Internal Medicine Ann Intern Med 147.11 (2007): 783. Web.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

Not applicable

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

Not applicable

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Hypertension is a silent killer and affects 1 billion people worldwide. Globally about 9 million people are killed by the potentially avoidable consequences of hypertension (WHO 2013). One in 3 adults over the age of 18 suffer from hypertension in the US and almost half of them do not have adequate control of their blood pressure to less than 140/90mm Hg (Nwankwo 2013). Almost 2/3rd of the patients with hypertension are therefore at increased risk of cardiovascular events (Elliot 2003). The socioeconomic burden of preventable conditions such as stroke, coronary heart disease, heart failure, and end-stage renal disease can be reduced by adequate control of hypertension.

Preventing hypertension and its complications is less expensive than interventions like CABG or dialysis if hypertension is not controlled (Elliot 2003). The Eighth Joint National Committee (JNC 8) published in JAMA in 2014, issued evidence based guidelines stating that, "Hypertension remains one of the most important preventable contributors to disease and death. Abundant evidence from randomized controlled trials (RCTs) has shown benefit of antihypertensive drug treatment in reducing important health outcomes in persons with hypertension."(James 2014)

Despite advances in antihypertensive drug therapy and the availability of guidelines, hospitalizations for hypertension continue to be a drain on health care resources. Hospitalization for hypertension is a PAC and a preventable ambulatory care sensitive condition that represents a failure of outpatient care system. A study analyzing National Hospital Discharge Survey (NHDS) conducted by the National Center for Health Statistics (NCHS) showed that 35,503 preventable hospitalizations occurred for hypertension between 1995 and 2010 in people over the age of 18 years. Age and sex standardized rates showed that blacks were 3 times more likely than whites to be hospitalized for hypertension and women had higher rates then men (Will 2013).

To improve accountability in the delivery of chronic care, AHRQ has developed a list of prevention quality indicators (PQIs) to identify ambulatory care sensitive conditions (ACSCs) and to measure rates of admissions that could have been potentially avoided with good outpatient care (AHRQ 2008). Even though hospitalizations for hypertension should be potentially avoidable in their own right; once they do occur, the index stay itself may have a potentially avoidable complication (PAC) or patients may develop a PAC during the post-discharge period. PACs lead to significant variability in outcomes including prolonged hospitalizations, readmissions and emergency room visits, all indicating poor outcomes that harm the patient, cause payers to incur unnecessary costs; and could be improved by providers (de Brantes 2011) (Yong 2010).
Therefore, there are many areas where improvement is possible in hypertension, making it a high priority aspect of health care. The PAC measures go beyond simple readmission rates and look for all-cause harms in patients with hypertension.

1c.4. Citations for data demonstrating high priority provided in 1a.3

1) World Health Organization. "A Global Brief on Hypertension: Silent Killer, Global Public Health Crisis." World Health Day 2013 (2013): WHO.int. WHO, 2013. Web. http://www.who.int/cardiovascular_diseases/publications/global_brief_hypertension.pdf

2) Nwankwo, Tatiana, MS, et al. "NCHS Data Brief - No. 133 - October 2013 U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES Centers for Disease Control and Prevention National Center for Health Statistics Hypertension Among Adults in the United States: National Health and Nutrition Examination Survey, 2011–2012." NCHS Data Brief 133 (2013): US Department of Health and Human Services: CDC. Web.

3) Elliot, William J. MD, PhD. "The Economic Impact of Hypertension" The Journal of Clinical Hypertension 5.3 (2003): 3-13.

4) James, Paul A., et al. "2014 Evidence-Based Guideline for the Management of High Blood Pressure in Adults." JAMA 311.5 (2014): 507-20. Web.

5) Will JC, Yoon PW. "Preventable Hospitalizations for Hypertension: Establishing a Baseline for Monitoring Racial Differences in Rates." Prev Chronic Dis 10 (2013): 120-165

6) Department of Health and Human Services, Agency for Healthcare Research and Quality. 2008. ''AHRQ Quality Indicators. Prevention Quality Indicators: Technical Specifications, Version 3.2''

7) de Brantes F, Rastogi A, and Sorensen CM. "Episode of Care Analysis Reveals Sources of Variation in Costs." Am J Manag Care 17.10 (2011): e383-e392

8) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. "The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010." Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable

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 subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

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

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

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Healthcare Associated Infections, Safety : Medication Safety, Safety : Readmissions, Safety : Venous Thromboembolism

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.)*
http://www.hci3.org/ecr_descriptions/ecr_description.php?version=5.2.006&name=HTN&submit=Submit

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_HTN_all_codes_risk_adjustment_06.30.15-635719646537139552.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Outcome: Number of patients who triggered an episode of hypertension (HTN), are followed for at least one-year, and had one or more potentially avoidable complications (PACs) during the episode time window.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The time window is the most recent 12 months of the episode, once a patient has triggered a HTN episode.

S.6. **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, 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.
Patients that have triggered a HTN episode, and are identified as having services for potentially avoidable complications (PACs), during the most recent 12 months of the episode. The enclosed excel workbook entitled NQF_HTN_all_codes_risk_adjustment_06.30.15 gives the detailed codes for PACs in the tabs entitled PACs I-9 and PACs I-10. PACs are identified only based on diagnosis codes.

Services for PACs are identified as follows:
a. Any service (professional, outpatient facility, ancillary) that is relevant to HTN and has a PAC code in any position on the claim
b. Any admission to an acute care facility, that is relevant to HTN
c. Any admission to a post-acute care facility that is relevant to HTN and has a PAC code in any position on the claim

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Adult patients aged 18 years and above who triggered an episode of hypertension (HTN) and are followed for at least one-year.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Please refer to the enclosed excel workbook entitled
NQF_HTN_all_codes_risk_adjustment 06.30.15

The target population is identified using the following criteria:
1. Using administrative claims database, patients with HTN are identified using one of two of the following criteria:
a. Patients having an office visit with a trigger code of HTN in any position, followed by a second confirmatory office visit (with a trigger code of HTN in any position), at least 30 days apart.
b. Patients with a Principal Dx of a HTN trigger code on an in-hospital stay claim.
The trigger codes for HTN are provided in the tab called "Triggers I-9" or "Triggers I-10".
2. The patient should have continuous enrollment for the entire time window with no more than 30 days as an enrollment gap, with the entity providing the data (so we can ensure that the database has captured most of the claims for the patient during the episode time window).
3. The patient should have a complete episode time window in the claims data – so there are at least 12 months of claims in the database for the patient.
4. Patient should be at least 18 years of age

Once the episode is triggered all relevant claims are assigned to the episode. Relevant claims could be inpatient facility claims, outpatient facility claims, professional services, laboratory services, imaging services, ancillary claims, home health, durable medical equipment as well as pharmacy claims across the entire continuum of care centered around the patient's episode of care. Relevant claims are identified as those that have a diagnosis code that matches the codes in the typical Dx codes tabs (Typical Dx I-9 or Typical Dx I-10), or in the PAC Dx codes tab (PACs I-9 or PACs I-10) AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant admissions to acute and post-acute care facilities are also included in the denominator.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the HTN measure if they are considered not relevant to HTN care.

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Denominator exclusions include exclusions of "patients" as well as "claims" not relevant to HTN care. Please refer to the enclosed excel workbook entitled (NQF_HTN_all_codes_risk_adjustment 06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
d. If the patient does not have at least 12 months of claims in the database (this helps eliminate incomplete episodes).
e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure if they meet one of the following criteria:
a. If none of the diagnosis codes on the claim are on the list of "triggers" or relevant diagnosis codes (either typical Dx or PAC Dx) for HTN.
b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for HTN.
c. The "principal" diagnosis on an inpatient stay claim during the episode time window triggers its own episode
d. The procedure code on a claim during the episode time window triggers its own episode

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
None

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*

Conceptual Model
Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, the remaining variance in PACs are due to factors that could be controlled by all providers that are managing or co-managing the patient, both during and after hospitalization.

Statistical Method:
Logistic Regression model to determine the probability of a patient incurring a PAC
Demographic variables, comorbid conditions, as well as clinical severity indicators are fed as independent risk factors into the model. Risk Factors are collected historically. Subtype information is collected from the index claim and any look-back period, if relevant. Subtypes are clinical severity indicators suggesting severity of the episode itself, for example, hypertensive heart disease in a HTN patient. For each patient the "predicted" coefficients from the risk adjustment models are summed to give the predicted

probabilities of the occurrence of a PAC.

Risk Factors :(Please refer to the enclosed excel workbook entitled (NQF_HTN_all_codes_risk_adjustment 06.30.15.xls). The risk factors along with their codes are listed in the tabs called “All Risk Factors I-9” and “All Risk Factors I-10” and also listed below:

AGE CONTINUOUS VARIABLE
GENDER FEMALE = 1 (MALE IS REFERENCE = 0)

Risk Factor #	Risk Factor Name
RF0101	Anoxic Brain Damage, persistent vegetative state
RF0102	Delirium, Meningitis, Encephalitis
RF0103	Previous Stroke, Paralysis
RF0104	Cerebral Palsy and Other Paralytic Syndromes
RF0105	Spinal Cord Disorders/Injuries
RF0106	Polyneuropathy
RF0107	Multiple Sclerosis
RF0108	Convulsions, Epilepsy
RF0109	Dementia
RF0110	Parkinson´s and Huntington´s Diseases
RF0111	Cerebrovascular Disease
RF0115	after care, rehabilitation
RF0201	visual loss, blindness, retinal tear, detachment
RF0301	ENT, Upper Respiratory Problems
RF0401	Respiratory Failure, O2, ventilator dependence
RF0402	Advanced COPD, Asthma
RF0403	Empyema, bronchiectasis, Pneumonias
RF0404	Aspiration Pneumonia, Laryngeal Problems
RF0406	TB, Pneumoconiosis, Aspergillosis
RF0407	Tobacco use, Lung disease due to External Fumes
RF0408	Other Lung Disease
RF0501	Previous Shock, Syncope, Vent Fibrillation
RF0503	Advanced CHF
RF0504	Cardiomyopathy, valve disorders
RF0505	Cardiac Arrhythmias, Heart Block
RF0506	Pacemaker, AICD
RF0507	Endocarditis, Other post surgical cardiac problems
RF0508	Other Cardiovascular Disease
RF0511	DVT, Pulm Embolism, Pulm Heart Disease
RF0512	Unstable Angina
RF0513	Hypotension, chronic, orthostatic
RF0514	Hyperlipidemia
RF0515	Intraaortic Balloon Pump
RF0516	ventricular assist device, ecmo, prolonged bypass
RF0517	Previous electrophysiology studies, cryoablation
RF0518	Recent AMI
RF0519	Previous PCI
RF0520	Previous CABG
RF0521	Previous Heart & Valve Surgery
RF0522	Previous aortic reconstruction
RF0523	Previos carotid endarterectomy
RF0524	Aortic and peripheral vascular disease
RF0525	Advanced Aortic and Vascular Disease
RF0601	GI Bleed

RF0602 Intestinal Obstruction/Perforation
RF0603 Acute Gastritis, Duodenitis
RF0604 Gastroduodenal Ulcer
RF0606 Intestinal Uro-genital Fistula
RF0607 Abdominal hernia w complications
RF0608 Vascular insufficiency of intestine
RF0609 Inflammatory Bowel Disease
RF0610 Irritable Bowel
RF0611 Diverticulitis, Meckel´s
RF0612 Digestive congenital anomalies
RF0613 Intestinal infection
RF0614 Esophageal Perforation, Hmg, Barretts, Compl Hiatal Hernia
RF0615 Abnormal weight loss
RF0616 Achalasia, Esophageal spasm, Stricture, Dysphagia
RF0617 GERD, Hiatal Hernia, Other Upper GI Disorders
RF0618 Previous Bariatric Surgery
RF0619 Hx of colon polyps, family Hx of colon cancer
RF0620 Enterostomy, GI devices, lap band
RF0701 Pancreatic Disease
RF0702 Perforation, fistula GB, bile duct, pancreas
RF0703 Gall stones, cholecystitis
RF0704 End-Stage Liver Disease
RF0705 Hepatitis, Cirrhosis, Other Hepatbiliary Disorders
RF0706 Recent Gall Bladder, Hepatobilary Surgery
RF0707 Acute Pancreatitis, pseudo cyst
RF0801 Bone/Joint/Muscle Infections/Necrosis
RF0802 Muscular Dystrophy
RF0803 Osteoporosis, ostetits deformans, pathological fracture
RF0804 Rheumatoid Arthritis and Inflammatory Connective Tissue Disease
RF0805 Gout and other crystal arthropathies
RF0806 Other arthropathies
RF0807 Osteoarthritis
RF0808 Joint Deformities
RF0809 Knee derangements
RF0810 Traumatic Dislocation Knee
RF0811 Dislocation Hip
RF0812 Synovitis, Ruture Tendon
RF0813 Status Knee Replacement
RF0814 Status Total Hip Replacement
RF0901 Decubitus Ulcer
RF0902 Skin and wound problems
RF1001 Diabetes, poor control
RF1002 Advanced diabetes
RF1003 diabetes
RF1101 Acute renal failure
RF1102 Dialysis Dependent
RF1103 Nephritis
RF1104 Chronic renal failure
RF1105 Urinary Tract Infections
RF1301 Endometriosis
RF1302 Fibroid uterus, benign tumors of female organs
RF1303 Pelvic Inflammatory disease
RF1304 Uterine prolapse, cystocele, vaginocele
RF1305 Female Harmonal Disorders

RF1306 Ovarian, Broad Ligament Disorders
RF1308 Other disorders of uterus, cervix
RF1309 Menopausal Disorders
RF1310 Menstrual Disorders
RF1401 Multiparity, multigravida
RF1402 Elderly Primi, other
RF1403 Poor obstetric history
RF1406 Cervical incompetence
RF1407 Abnormalities of uterus, female genital tract
RF1408 Hypertension, pre-eclampsia in Pregnancy
RF1409 Severe pre-eclampsia w HTN, Eclampsia
RF1410 Maternal, gestational diabetes, large for date
RF1411 Genital Herpes
RF1412 Infections of genitourinary tract, venereal disease in pregnancy
RF1413 Infectious Diseases in Mother
RF1414 Cardiovascular disease in Mother
RF1415 Mental Disorders in Mother
RF1416 Epilepsy in Mother
RF1417 Liver and biliary tract disorders in mother
RF1418 Kidney Disease in Mother
RF1419 Other Maternal conditions
RF1421 Cephalopelvic Disproportion due to maternal causes
RF1436 Peripartum Cardiomyopathy
RF1441 Previous Cesarean section
RF1450 Maternal Obesity, previous Bariatric Surgery
RF1454 Previous Rupture Uterus, Obstetrical Trauma
RF1458 Complicated Pregnancy Delivery
RF1460 Thrombophlebitis, DVT during Pregnancy
RF1461 Puerperal Sepsis, other major puerperal complications
RF1462 Obstetrical Embolism, Air, Amniotic Fluid, Pulm, Pyemic
RF1467 Tobacco Use in Mother
RF1601 Bleeding Disorders
RF1602 Severe Hematological Disorders
RF1603 Disorders of Immunity
RF1604 Nutritional and other Anemias
RF1605 Long-term use of anticoag, Aspirin
RF1701 Head and Neck Cancers
RF1702 Lung and Intrathoracic Cancers
RF1703 Neuroendocrine, Myeloproliferative Cancers
RF1704 Poorly differentiated, Secondary, Metastatic Cancers
RF1705 Other Tumors
RF1706 Acute Leukemia
RF1707 Cancer uterus, localized female organs
RF1708 Colorectal, Hepatobiliary and other GI cancers
RF1709 Breast, Prostate, Thyroid cancers
RF1710 Testicular Cancer and localized of male organs
RF1711 Cancer of Bladder and Urinary Tract
RF1712 Musculoskeletal Cancers
RF1801 Sepsis, MRSA, Opportunitistic infections
RF1901 Schizophrenia
RF1902 Major Depressive, Bipolar, and Paranoid Disorders
RF2001 Drug/Alcohol Psychosis
RF2002 Drug/Alcohol Dependence
RF2101 Drug Reactions, long term use of drugs

RF2102 Intra-abdominal injury
RF2201 Extensive Third-Degree Burns
RF2301 Major Organ Transplant Status
RF2302 Artificial Openings for Feeding or Elimination
RF2303 Complications of Medical & Surgical Care and Trauma
RF2304 severe morbid obesity
RF2305 morbid obesity
RF2306 obesity
RF2307 mild sleep apnea, hypoventilation
RF2308 moderate sleep apnea, hypoventilation
RF2309 obstructive sleep apnea
RF2310 Severe Protein-Calorie Malnutrition
RF2311 Mild-mod malnutrition
RF2401 Severe Head Injury
RF2402 Major Head Injury
RF2403 Vertebral Fractures without Spinal Cord Injury
RF2404 Falls, Fractures
RF2405 Amputation
RF2501 HIV/AIDS

Subtypes for HTN
Hypertensive Heart Disease
Renovascular and other secondary hypertension

The prevalence of the risk factors in our reference dataset are listed in the enclosed workbook entitled NQF_HTN_all_codes_risk_adjustment 06.30.15.xls – see tab "Risk Factor Prevalence". The output of the regression model are given in the same workbook in the tab "Risk Model".

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Please refer to the enclosed excel workbook entitled (NQF_HTN_all_codes_risk_adjustment 06.30.15.xls).

Assembling the Denominator:

Using administrative claims database, patients with HTN are identified using one of two criteria: 1) Patients having an office visit with a trigger code of HTN in any position, followed by a second confirmatory office visit (with a trigger code of HTN in any position),

at least 30 days apart, 2) Patients a Principal Dx of a HTN trigger code on an in-hospital stay claim. The trigger codes for HTN are provided in the tab called "Triggers I-9" or "Triggers I-10".

Patients are retained if they are 18 years of age or more, do not have a missing gender, have at least 12 month of claims in the database, have a maximum of 30-day enrollment gap for the entire episode time window, and have no outlier episode costs. All relevant professional, laboratory, imaging, ancillary and other claims that are incurred during the episode time window are included as part of the episode. Claims are considered relevant to HTN care if they have one of the diagnosis codes, as listed on the tab entitled Triggers I-9, Triggers 1-10, PACs I-9, PACs I-10, Typical Dx I-9, or Typical Dx I-10 in any position on the claim AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant admissions to acute and post-acute care facilities are also included in the denominator. All relevant pharmacy claims carrying codes that match the ingredients listed in the Pharmacy tab of the enclosed workbook are also included as part of the episode.

If a patient has more than one concurrent episode, and the claim is relevant to both episodes, the claim could get multi-assigned, except in the case of procedural episodes that get carved out with respect to the index stay. So if an inpatient stay claim carried a principal Dx code that matched the trigger diagnosis code for HTN but they also had a procedure code for CABG (coronary artery bypass surgery), the stay claim would get uniquely assigned to CABG and not be counted with HTN.

Once all the episodes are assembled, episodes that match the exclusion criteria, such as those with outlier costs, are flagged (those with total episode costs less than 1st percentile or greater than 99th percentile), and excluded from the final analysis.

Assembling the Numerator:

For every episode included in the denominator, services are flagged as having a PAC (potentially avoidable complication) based on the criteria listed below:

Any service (professional, outpatient facility, ancillary) with a PAC code in any position on the claim
Any admission to an acute care facility, that is relevant to HTN

Relevant claims that do not have any PAC codes, and do not qualify as a PAC based on the criteria outlined above, are listed as typical claims. All included relevant pharmacy services are flagged as typical. Patients that have even a single PAC claim are counted as part of the numerator.

Calculating the measure:

Proportion of HTN patients that have PACs is simply the ratio of patients with PACs within the HTN population and is called the PAC rate as shown in the equation below:

PAC rate = Patients with HTN that have at least one PAC claim / Total number of HTN patients

A flow chart demonstrating the series of steps and the counts of patients at each step is shown in tab entitled Decision Tree of the enclosed workbook called NQF_HTN_all_codes_risk_adjustment 06.30.15.xls

Drill Down Calculations:

Further analysis from this construct helps create actionable reports.

For example as shown in the tab labeled PAC overview, not only do we have the PAC rate for a population, we can break them down by the PAC type – type 1 being directly related to HTN and so actionable by the servicing physician, while type 2 PACs are related to patient safety failures and can be improved by process improvement. Additionally, analyzing potentially avoidable admissions helps focus strategies in reducing them.

Risk Adjustment:

Once we have the observed PAC rates, we risk-adjust them for patient factors such as patient demographics, comorbidities collected historically, and for severity of illness or procedure using subtypes collected from the index stay and / or look-back period. This helps adjust for factors outside the providers control and levels the playing field for provider performance comparisons.

Unit of Analysis:
The unit of analysis is the individual episode.

Dependent Variable:
The dependent variable is a dichotomous variable indicating whether an episode had one or more claims assigned as a PAC (=1) or not (=0).

Independent Variables:
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_HTN_all_codes_risk_adjustment 06.30.15.xls

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., unstable anginahypertensive heart disease). Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_HTN_all_codes_risk_adjustment 06.30.15.xls

As mentioned previously, to avoid creating perverse incentives all comorbidities and subtypes are identified prior to or at the very start of the episode. None are identified during the episode period.

Statistical Methods
We use logistic regression to model the probability of at least one PAC occurring during the episode. Only comorbidities and subtypes are included in the models as covariates if they are present in at least 10 episodes to prevent unstable coefficients. No further model building is conducted after the initial models are built. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but it does not make it a priority that all covariates in the model be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity of a episode condition, and lets each regression model determine for itself which of the factors are more significant for a specific episode. Non-significant covariates in episode models can not overly influence predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

When more than one line of business is included in the data, separate models are calculated for each sample (i.e., commercial, Medicaid etc.).

Provider Attribution and calculating PAC rates by provider:

Once episodes are constructed they are attributed to providers based on one of the various attribution rules. For HTN, episodes are attributed to the primary care physician, internist or other physicians cardiologist with the highest count of office visits.

Using the logistic regression technique described above, a model is developed that gives estimates for each risk factor and subtype for the patients in the population analyzed. These estimates are used to develop patient-level probabilities for the occurrence of PACs. The patient-level probability estimates are summed to construct aggregated measures (e.g., facility/provider-level). This method is similar to the methods employed by the Centers for Medicare and Medicaid Services (CMS) and endorsed by the National Quality Forum (NQF) to construct similar facility- and practice-level measures (i.e., mortality, readmissions, etc.):

For each provider, the number of actual observed occurrences of the outcome is summed across all attributed patients with that episode, to give the observed PAC rates for the provider.

Similarly adjusted probabilities from the risk adjustment models are summed across all attributed patients to give expected

PACs for the provider.

The observed sum is then divided by the summed probabilities (O/E). This number yields whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). This calculation yields a practice-level unstandardized performance ratio.

To facilitate accurate comparisons of rates across units of analysis, this ratio is then standardized to the community rate using the indirect method. Specifically, the provider-level rate is multiplied by the expected community rate, calculated as the sum of adjusted probabilities for every individual in the sample across all providers in the analysis. This measure, known as the standardized rate, represents what the unit's risk-adjusted rate would be for the outcome of interest if its patient population was reflective of the of the overall community.

The formula for this calculation is as follows:

Adj Outcome_j={(SUM Observed_ij)/(SUM Prob_ij)} × {(SUM Prob_i) / (# of episodes)}
Where individual is attributed to unit of analysis j (e.g., practice, provider, etc.)

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If patient related data is missing, the case is deleted from both the numerator and the denominator.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The information is based on a two-year claims database from a large regional commercial insurer. The database has over 3.2 million covered lives and $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

The methodology can be used on any claims database with at least two years of data and a minimum of 150 patients with the index condition or hospitalization. Having pharmacy data adds to the richness of the risk-adjustment models.
The calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.
The methodology has been tested on databases of several health plans as well as on a few employer databases.

No data collection instrument was used.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Team, Health Plan, Population : National, Population : Regional, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinician Office/Clinic, Other
If other: Across the care continuum

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2748_HTN_Testing_Reliability_Validity_HCI3-635719660723458859.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2748
Measure Title: Proportion of Patient with Hypertension (HTN) that have a Potentially Avoidable Complication (during the episode time window)
Date of Submission: 06/30/15
Composite Construction:

☐Two or more individual performance measure scores combined into one score
☐ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)
☐ Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.
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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. 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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g.,*

Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

The information is based on a two-year administrative claims database from a large regional commercial insurer. The database contains medical and pharmacy claims on over 3.2 million covered lives and more than $25.9 billion in "allowed amounts" for costs.

1.3. What are the dates of the data used in testing? April 1, 2012 – December 17, 2014

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.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Integrated Delivery System	☐ other:

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

There were a total of 23,126 providers in the data set. Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the reliability calculations. After this exclusion, there were 3,702 providers left.

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

After exclusions (see 2b.3.1 below), there were a total of 262,273 episodes of HTN were included in the testing and analysis. Patients in these episodes were, on average, 53.0 years of age (range 18-64) and 46% were female. We did not have race information on these patients. All patients for this analysis had a trigger inpatient claim of HTN as identified in our code tables.

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.

For the reliability analysis, we restricted the data to only providers with at least 10 attributed episodes. For risk adjustment, all episodes were used in the analysis, regardless of the provider to which they were attributed.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

None of the analyses included SDS variables.

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*)

We assessed the reliability of the measure to demonstrate that it sufficiently differentiates performance between providers using the beta-binomial method, which is applicable to measures of this type. Reliability is a measure that distinguishes the signal (the extent of performance variation between entities that is due to true differences in performance) from statistical noise. Our approach follows directly from the methods outlined in the technical report "The Reliability of Provider Profiling: A Tutorial" by J.L. Adams.

Reference:

Adams JL. The Reliability of Provider Profiling: A Tutorial. Rand Corporation. http://www.rand.org/pubs/technical_reports/TR653.html.

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*)

The table below provides a summary of the reliability score for different minimum sample size thresholds. For complete results, refer to the workbook entitled, NQF_HTN_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab to see provider-specific results.

Reliability Scores	**Minimum # Episodes Per Provider**		
	>=10	**>=25**	**>=50**
# of Providers (%)	3,702 (100)	2,011 (54)	1,039 (28)
Median (IQR)	0.79 (0.67, 0.89)	0.87 (0.81, 0.92)	0.92 (0.89, 0.95)
Range	0.49-1.00	0.71-1.00	0.83-1.00

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

Reliability scores can vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or variation across patients within providers) whereas a

reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (see Adams, 2009 cited above).

Although some scores among providers with at least 10 episodes were low, many had scores that met or exceeded the minimum acceptable level for reliability. Moreover, limiting providers to those with at least 25 or 50 episodes, scores were consistently good. These results demonstrate that the measure sufficiently differentiates providers' performance.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b2. VALIDITY TESTING

Note: *Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

2b2.1. What level of validity testing was conducted?

☐ **Composite performance measure score**

- ☐ **Empirical validity testing**
- ☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)
- ☐ **Systematic assessment of content validity**

☐ **Validity testing for component measures** *(check all that apply)*

Note: *applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

- ☐ **Endorsed (or submitted) as individual performance measures**
- ☐ **Critical data elements** (*data element validity must address ALL critical data elements*)
- ☐ **Empirical validity testing of the component measure score(s)**
- ☐ **Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. 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)*

Content validity was built into the development of the definitions of potentially avoidable complications (PACs). This involved working with clinicians who are experts in their respective fields and specific to the episodes for which PACs are being measured. In particular, the clinical experts focused on whether or not a potentially avoidable complication can be deemed as such for a specific episode of care, and help defined and review all of the diagnosis and procedure codes for each PAC. The enclosed link lists clinicians who have participated in the various Clinical Working Groups (http://www.hci3.org/content/clinical-working-group-contributors). Some of the clinical experts have also participated in monthly webinars that highlight the clinical aspects of these measures (http://www.hci3.org/content/using-ecrs-providers).

In addition, we illustrate that our measure has face validity in several ways.

Beyond the up front work performed by clinical experts, the validity of the measure has also been tested in various real world settings. For example, we have presented results of claims data analyses that reveal the frequency and costs of PACs to physicians in several different healthcare systems involved in our pilot site implementations, as well as to medical directors from the employer coalitions and the health plans that provided the dataset to run the analyses. Some of these implementations include the Pennsylvania Employee Benefits Trust Fund and local provider groups and hospital, Horizon Blue Cross Blue Shield of NJ and many physicians and health systems.

In addition, we have performed dozens of analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:

- Vermont Payment Reform Commission
- Maine Health Management Coalition
- WellPoint / Anthem CT
- NY State Medicaid
- CT Medicaid
- CO All-payer Claims Database, Center for Improving Value in Health Care

These analyses and their results have influenced, and continue to influence, the development of various public reporting, payment reform and delivery system reform efforts. To-date, we have never experienced either wholesale or partial rejection of the results of analyses showing rates of PACs, which demonstrates the level of acceptability – face validity – of the measures from the payer, policymaker, employer and payer communities.

As importantly, measures of potentially avoidable complications have face-validity with consumers. In a series of focus groups, Judy Hibbard and colleagues[1] examined the impact of presenting information about price and quality of certain providers in influencing the decisions of consumers. They tested the validity of PACs as a discriminator of quality, as well as other measures of quality, and used the dollar symbol to illustrate the level of price, much like is done for restaurant reviews. When the PAC measure was used, respondents selected the providers with the lowest PAC rates with a high level of confidence in choice, and used it as a surrogate for a

strong quality signal. To the contrary, when more standard measures of quality were used, consumers tended to ignore them and use price as a surrogate for quality. As such, what the researchers found is that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Finally, our measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs)[2], Hospital Inpatient Quality Reporting measures [3], Avoidable Readmissions [4,5], AHRQ defined patient safety indicators (PSIs) [6], NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates [7].

References:

[1] Hibbard JH, Greene J, Sofaer S, Fiminger K, and Hirsh J. An Experiment shows that a well-designed report on Costs and Quality can help consumers choose High-Value Health Care. *Health Affairs* 2012; 31(3): 560-568. doi: 10.1377/hlthaff.2011.1168

[2] CMS defined Hospital Acquired Conditions: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

[3] CMS operated Hospital Inpatient Quality Reporting Program: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

[4] Jencks SF, Williams MV, and Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. N Engl J Med 2009 (Apr); 360 (14): 1418-1428. doi: 10.1056/NEJMsa0803563.

[5] Casalino LP, Pesko MF, Ryan AM et.al. Small Primary Care Physician Practices have low rates of Preventable Hospital Admissions. Health Affairs, 2014; 33(9): 1-9. doi: 10.1377/hlthaff.2014.0434.

[6] Agency of Healthcare and Quality defined Patient Safety indicators: http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx

[7] NQF endorsed measures: Quality Positioning System: http://bit.ly/1E5ZdP7

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Not applicable.

2b2.4. 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?*)

Given the significant clinical input that went into developing the measure, the widespread use and acceptance the measure has gained among a wide variety of individuals and organizations across the health system (public and private payers, clinicians, consultants, patients, etc.) [1-13], and the parallels between this measure and other measures that are in widespread use, this demonstrates that the measure has strong face validity.

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)

2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.

3. de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.

4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.

5. Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.

6. Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.

7. François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)

8. de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

9. Rastogi A, Mohr BA, Williams JO, Soobader MJ, de Brantes F. Prometheus Payment Model: Application to Hip and Knee Replacement Surgery. Clin Orthop Relat Res 2009; 467(10): 2587-2597.

10. de Brantes F and Rastogi A. Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. The Commonwealth Fund 40, publ. 2008; 1146:1-14.

11. de Brantes F, Gosfield A, Emery D, Rastogi A and G. D'Andrea, "Sustaining the Medical Home: How Prometheus Payment Can Revitalize Primary Care", Robert Wood Johnson Foundation Report, May 2009, http://www.rwjf.org/pr/product.jsp?id=42555, accessed October 2009.

12. de Brantes F, Camillus J. Evidence-informed case rates: a new health care payment model [Internet]. New York (NY): Commonwealth Fund; 2007 Apr [cited 2007 May 20]. Available from: http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=478278, Accessed Aug 1 2013.

13. Satin DJ, and Miles J. Performance Based Bundled Payments: Potential Benefits and Burdens. Available from: http://student.med.umn.edu/p4p-new/sites/default/files/MN%20Med%20Bundles%20Special%20Report%20-%20Satin.pdf, Accessed Aug 1 2013.

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

No formal exclusion testing was done since no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers.

Exclusions included exclusions of "patients" as well as "claims" not relevant to HTN care. Please refer to the enclosed excel workbook entitled (NQF_HTN_all_codes_risk_adjustment_06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
d. If the patient does not have at least 12 months of claims in the database (this helps eliminate incomplete episodes).
e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. “Claims” are excluded from the measure if they meet one of the following criteria:
a. If none of the diagnosis codes on the claim are on the list of “triggers” or relevant diagnosis codes (either typical Dx or PAC Dx) for HTN.
b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for HTN.
c. The “principal” diagnosis on an inpatient stay claim during the episode time window triggers its own episode
d. The procedure code on a claim during the episode time window triggers its own episode.

2b3.2. 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*)

We started with a total HTN population of 409,442 episodes. After all the exclusions were applied, the remaining HTN population included in the analysis consisted of 262,273 episodes. As mentioned above, no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers. As such, no formal exclusion testing was done.

2b3.3. 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*)

No formal analysis was done on the impact of exclusions on performance scores.

Descriptive Explanation:

Exclusions of patients were for the following reasons. Some are for comparative purposes and some for medical reasons.

(a) Comparative Purposes:
We excluded patients that did not have complete enrollment for the entire episode time window. This was done to ensure that the database had complete information on patients to be able to create the entire episode. Including patients with only a partial episode window could distort the measure by artificially reducing the actual count of patients with PACs.

(b) Medical Reasons:
Patients with outlier costs (less than 1st percentile value or greater than 99th percentile) were considered to be different from the general pool, and excluded from both the numerator and the denominator. This is another way to ensure that episodes are complete (because incomplete episodes may have very low costs), and do not bring in random noise into the analysis due to inappropriate codes or services (high outliers).

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1./S13 What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with 170 potential risk factors** and episode specific subtypes
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.1.2. If an outcome or resource use 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.

2b4.2/S14. Identify the statistical risk model variables (Name the statistical method – e.g., logistic regression and list all the risk factor variables.
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1)

identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_HTN_all_codes_risk_adjustment 06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., Hypertensive Heart Disease, Renovascular and other secondary hypertension). Subtypes are specific to each unique episode and are included in the models only if they are present at the start of the episode. Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_HTN_all_codes_risk_adjustment_06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Candidate comorbidities and subtypes were included in the models as covariates if they were present in at least 10 episodes to prevent unstable coefficients.

2b4.2.1/S15. Detailed risk model specifications including coefficients, equations, codes with descriptors, definitions(may be attached in an Excel or cvs file)
All Risk Factors with their coefficients are detailed in the enclosed workbook called NQF_HTN_all_codes_risk_adjustment_06.30.15.xls – Please reference the tabs titled Risk Factor Prevalence and Risk Model.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Risk factors are comorbidity indicators collected from historical claims before the start of an episode. These are universally applied across all episodes. This list was selected based on input from clinical experts in clinical working groups. In addition, the Clinical Working Groups identified episode specific severity markers that were called episode subtypes and they help distinguish an episode as being more severe than another.

All risk factors and subtypes must be present prior to, or at the start of the episode and are identified using diagnosis codes in the patient's historical claims.

To be included in the risk adjustment models, any risk factor or subtype must be present in at least 10 episodes. Beyond this no further model building was conducted to add or remove risk factors or subtypes from the model after it was initially run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates cannot overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
As explained above, no formal analysis was conducted to select risk factors. In fact, all potential risk factors and subtypes with a count of at least 10 episodes were retained to serve as predictors. The goal was to achieve a more complete explanatory model rather than achieve parsimony.

Please reference the tabs titled Risk Model in the NQF_HTN_all_codes_risk_adjustment_06.30.15.xls workbook to see the list of risk factors that met the selection criteria.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Not Applicable since our analysis did include SDS variables

2b4.5. 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*)

Model Development Approach

We used logistic regression to model the probability of at least one PAC occurring during the episode. The model included all covariates that were identified through the process above. No further model building was conducted after the initial model was run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates can not overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

For a more complete description of the risk adjustment approach, please see the document entitled, "PACs and Severity Adjustment Fact Sheet" that accompanies this submission.

Approach to Model Testing and Validation

To determine the validity and performance of the model, we used the split sample method to divide the patient sample randomly into: 1) the model building data set (80% of the sample) and 2) the test data set (20% of sample. The model was built using logistic regression on the first data set and then the coefficients from the development model were tested in the second dataset. Area under the curve (AUC) and the c-statistic were used to compare the predictive ability of the model in each of the data sets. Hosmer-Lemeshow Goodness-of-Fit tests and comparisons of observed to expected probabilities across risk deciles were further examined to assess the model's overall predictive accuracy.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

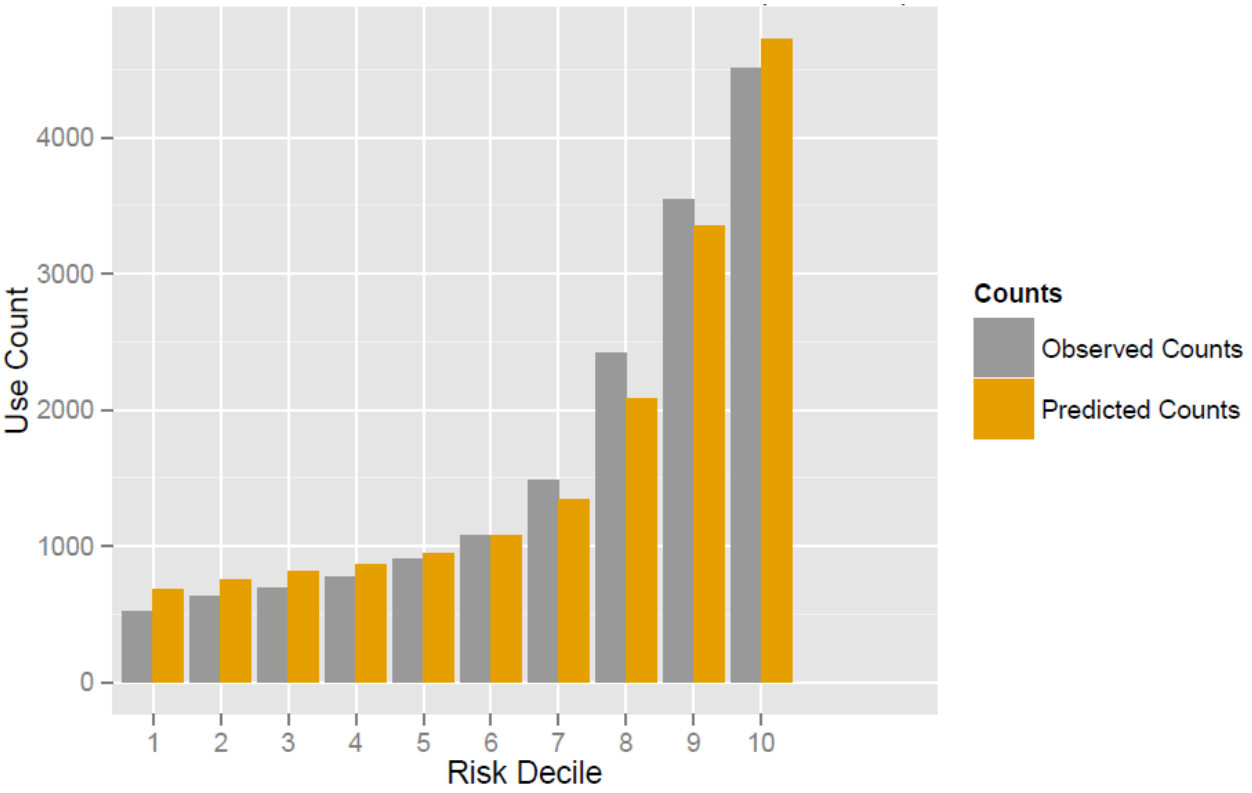
Sample	Accuracy (%)*	AUC
Test	79.0%	0.800
Validation	79.2%	0.801

*Episodes with predicted probabilities <50% were classified as having a predicted 0 (not having a PAC). Episodes with predicted probabilities >50% were classified as having a predicted 1 (having a PAC)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Sample	Chi Square	Degrees of Freedom	p-value
Test	1500.1	8	<0.0001
Validation	337.7	8	<0.0001

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:



2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The C statistic is a measure of the extent to which a statistical model is able to discriminate between a patient with and without an outcome. The c-statistic ranges from 0.5 to 1.0. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong [1]. The purpose of the model is to adjust for patient-related factors. The remaining unexplained differences in PAC rates are due to factors that could be controlled by all providers that are managing or co-managing the patient, during the entire episode time window.

The c-statistics of the testing and validation samples (0.800 and 0.801, respectively) indicate that the risk models have strong discriminatory power. Indeed, the accuracy values show that the model correctly predicts whether an episode had or did not have a PAC just under 80% of the time, well above what would be expected if the predictions were made at random (i.e., 50%). Although the H-L tests were significant, meaning that the model is not a good fit, test is generally known to be sensitive to the number of groupings used and sample sizes. As shown by the risk decile plot, however, the model predicts PACs similar to the number of observed PACs across each of the deciles.

Overall, the results suggest that the model has sufficient predictive power.

Reference:
[1] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Edition)*. New York, NY: John Wiley & Sons; 2000.

2b4.11. 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*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To directly compare PAC rates across providers while also appropriately accounting for differences in patient severity, we calculated a risk-standardized PAC rate for each provider. This method is similar to calculations used by others for reporting outcomes measures [1]. For each provider, the ratio of observed attributed episodes with PACs to the expected number of attributed episodes with PACs given the patient's risk factor and estimated from the risk-adjustment model was calculated. This number yielded whether the provider had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). We then multiplied this ratio by the overall expected PAC rate across all providers to obtain the risk-standardized PAC rate for the provider. This measure represents what a provider's PAC rate would be if its patient population was reflective of the overall population.

Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. Comparison of risk-adjusted PAC rates gives a measure of the provider's relative performance. Our analysis compared risk-standardized PAC rates across providers. We analyzed various descriptive statistics including the range in PAC rates, medians, interquartile range, etc.

Reference:

[1] See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

2b5.2. 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*)

Summary of Unadjusted and Adjusted Performance Scores Across Providers:

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25

Unadjusted		
Median (IQR)	30% (20%, 43%)	29% (20%, 41%)
Range	0%-100%	3%-100%
Adjusted (RSPR)*		
Median (IQR)	31% (24%, 38%)	31% (25%, 36%)
Range	0%-173%	4%-112%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_HTN_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

2b5.3. 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?*)
Even after right-adjustment, the variation in risk-adjusted rates suggests there are meaningful differences in performance between providers in risk-standardized PAC rates for patients with an episode of HTN.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This criterion is directed 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 SDS 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.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

2b6.2. 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*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

If patient related data is missing, the entire patient is excluded from the numerator as well as the denominator.

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence.

According to our measure definition, in constructing the measure it is possible for a provider to have only one or some types of PACs and not others. Alternatively, the provider may have all PAC types occur for their patients. The measure only considers whether any PAC occurred regardless of the type, and all PAC types are weighted equally, therefore we believe, there is no potential for the absence of specific PAC types to bias performance scores for individual providers.

For these reasons, no formal analyses were done on missing data.

2b7.2. 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*)
Not applicable

2b7.3. 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*)

Not applicable

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook entitled NQF_HTN_all_codes_risk_adjustment_06.30.15.xls provides outputs from empirical analysis. The tab labeled "PAC overview" demonstrates percentage of episodes that had at least one PAC, and provides the breakdown of PACs: 1) by the type of PAC whether directly related to index condition or due to patient safety failures; 2) the setting of the PAC, whether seen in the in-patient setting, out-patient facility or during professional visits; and 3) preventable hospitalizations.

The "PAC Drill Down Graph" provides further detail on each component of the PAC and their frequency. As can be seen by the individual counts and the graph, while each individual PAC may have such small occurrences that no meaningful comparisons in provider performances could be made; together, they add value to provide a comprehensive picture that result in meaningful numbers. The aggregation of PACs to a comprehensive, composite measure, in itself provides the parsimony that is so desirable.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

All PACs, as clinically defined by the subject matter experts were used with equal weighting. Since the emphasis of the PAC measure is to identify the occurrence of PACs in any setting, a simple and straightforward approach was adopted.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc*.; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

No formal analysis was performed.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Additionally, if a patient had one or more PACs, it is simply counted as a 1, i.e., flagged as having a PAC. The measure only considers whether any PAC occurred regardless of the type, or the site, and all PAC types are weighted equally. Therefore, no formal analysis of individual components was performed.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

We chose not to weight the components of the measure.

Considerations were given to the fact that preventable hospitalizations may be given more weight, than PACs identified in a doctor's office. Similarly PACs in an in-patient setting may have more serious implications on a patient's ultimate outcome, than PACs occurring in an outpatient setting. Additionally, preventable hospitalizations as well as index hospitalizations, each with longer lengths of stay, may have serious PACs. But how do we weigh these effects? An alternative model was considered, where cost could be considered as a surrogate for the weights. Higher cost PACs could imply more serious PACs. However, differences in costs could be driven by many issues other than the PAC itself, such as unit price of the service, method of reimbursements, contracting arrangements etc.

Furthermore, in-patient facility billing does not allow for the distinction of PAC related costs from other costs within the stay. We would fail to capture PAC related costs within the stay and potentially underweight those. As a result, the decision was made to avoid weighting and keep the measure as a straightforward count.

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

Measuring all providers with the same yardstick will provide consistent results and reasonable comparisons over time. If the goal is to reduce PACs, then the PAC measure as was constructed with the help of various experts in the field would provide reasonable comparisons. A word of caution however pertains to the sample size of the provider panel before making any reasonable conclusions.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved for high reliability in one dataset would apply to another.

2d3. Empirical analysis demonstrating that the approach for handling missing data minimizes bias (*i.e., achieves scores that are an accurate reflection of quality*)**.**

Note: *Applies to the overall composite measure; the focus is on missing data rather than exclusions, which are considered in 2b3.*

Please refer to section 2b7

2d3.1. What is the overall frequency of missing data and the distribution of missing data across providers?

2d3.2. Describe the method used to compare approaches for handling missing data (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification*

2d3.3. What were the statistical results obtained from the analysis of missing data? (e.g., *results of sensitivity analysis of effect of various rules for missing data; if no empirical analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

2d3.4. What is your interpretation of the results in terms of demonstrating that the approach used for missing data minimizes bias? (i.*e., what do the results mean in terms of supporting the selected approach for missing data; if no empirical analysis, provide rationale for the selected approach for missing data*)

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

As part of our general implementation of these measures and related analyses, we have worked through dozens of different and sometimes very large datasets. From Medicare to Medicaid to regional and national commercial carriers, as well as individual employers, the principal lesson learned is the heterogeneity of the data sets and the significant variability in fill rate of critical data elements. As a result, we have created highly specific recommendations for which data elements are required to ensure measure validity, the accuracy of those data elements, and their completeness in the dataset. When claims datasets are organized in the way we specify in the measure analysis, and contain the coding information required, the analysis of the measure and its results are highly reliable.

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 calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.

We also plan on providing a limited automated analysis, at no cost, on our website.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Payment Program Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/ Blue Cross Blue Shield of New Jersey http://www.horizonblue.com/ Pennsylvania Employee Benefits Trust Fund https://www.pebtf.org/ Quality Improvement (Internal to the specific organization) Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/

4a.1. For each CURRENT use, checked above, provide:

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

Measures associated to potentially avoidable complications (PACs) are in use today with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system (public and private payers, clinicians, consultants, all-payer claims database stewards, etc.) [1-8]. They are being used in various capacities in different pilot site implementations. To name a few:

•BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
•BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers

In addition, the PAC measures are incorporated by the following organizations in their bundled payment programs:

•BCBSSC – for CABG and PCI programs
•Horizon BCBSNJ– for CHF and CABG programs
•BCBSNC
•PEBTF in PA

http://www.ajmc.com/interviews/Lili-Brillstein-on-How-Bundled-Payments-Are-Tranforming-Healthcare

In these programs they look at PACs related to the measure for process improvement activities and for practice re-engineering.

We have created reports for rates of PACs for the following organizations:
-Vermont Payment Reform

-Maine Health Management Coalition
-WellPoint / Anthem CT
-NY State Medicaid
-CT Medicaid
-CO All-payer Claims Database, Center for Improving Value in Health Care

There are several companies that are leveraging these measures to create analytics and software for customers – these include HealthQx, Aver Informatics, McKesson, and TriZetto.

Below are some references that highlight our work with Potentially Avoidable Complications (PACs).

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)
2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.
3. de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.
4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.
5. Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.
6. Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.
7. François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)
8. de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

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

4a.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.*)
Measures associated with PACs are currently in use as described in the prior section. In addition, we are working with several not-for-profit and for-profit organizations to provide them with the algorithms needed to calculate rates of potentially avoidable complications. Some of these organizations include:

Fair Health – based in NY and whose mission is to increase transparency of provider cost and quality,

CastLight – based in CA and serving large employers. We currently provide CastLight with Bridges To Excellence recognitions and will work with them to augment provider transparency by using PAC measures,

MA APCD (Massachusetts All Payers Claims Database) Council – we currently have an agreement in place with the MA APCD Council to produce PAC measures on hospitals and physicians and report back to the council with tests of reliability and validity of the measures. The purpose is to authorize the publication of these measures,
Maryland Health Care Cost Commission – we have a two year agreement to produce measures of cost and quality for public

dissemination.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

We do not have any public information to share about the improvements in rates of potentially avoidable complications, as the implementation of these measures is too recent to provide valid comparisons. Further, some of the definitions of PACs have changed since the measures were initially endorsed, making comparisons even more difficult and unreliable.

Nevertheless, the variation in performance scores presented in Section 1b.2 indicates that there are differences between providers in their risk-adjusted PAC rates (higher scores equal worse performance). This suggests that real opportunities exist to identify lower performing providers and reduce the overall occurrence of PACs.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Performance results provide summary PACs rates by provider, which can be used by payers and providers in a number of ways to improve the quality of care.

From the payer perspective, payers can use this information to 1) create a high-value provider networks, 2) work with high-value providers to share best practices, 3) incentivize low-value providers to improve, 4) modify their insurance design to activate consumers to select the right care from the right providers at the right time.

From the provider perspective, providers can 1) view services and activity for their patients longitudinally across the entire care continuum, such as frequency of readmissions and ED visits and drill down on patients with high PAC rates, 2) review actionable drill down reports to identify the most frequent PACs across all patients to create care pathways and process improvement plans to impact the most frequent PACs.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended consequences were reported, but there is the potential for:
1. Under-coding of PACs in the claim stream resulting in under-reporting the actual rate and/or providers gaming the measures
2. Payers calculating the measures even with inadequate sample sizes and using the results to penalize providers

The measure is designed for transparency efforts and to spur quality improvement. Detailed PAC reports can help providers identify areas of quality improvement. Even detailed reports of small samples of patients can be helpful for quality improvement purposes, but not for public reporting. To mitigate the potential for invalid provider comparisons, we specify in this submission the minimum sample size needed to ensure the reliability of a provider's score. Ultimately, there isn't any good way to prevent provider gaming of the measure by under-coding claims, however, under the current DRG payment methodology, many providers would be penalized by under-coding PACs since these codes often result in the assignment of more complicated DRGs.

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)

0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

5a. Harmonization

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 completely harmonized?

No

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

Some of the measures listed in the prior section are, fully harmonized with the submitted measure, in particular, 0705, 0708, and 0709. Other measures such as 0337 and 0450 are in fact, subsets of our measure. However, there are some measures that are not harmonized, in particular the 30-day all-cause readmission measures and the Hospital wide all-cause readmission measure. While the submitted PAC measure include hospitalizations and readmissions that occur during the episode time window, the hospitalizations, by definition, have to be relevant to the underlying condition. For chronic conditions, most relevant hospitalizations within the entire episode time window are considered potentially avoidable. PACs include readmissions and are designed to enable accountability at the locus of provider control as well as some shared accountability between settings, centered around a patient, and for a specific medical episode of care. In that sense, they are consistent with the all-cause 30-day readmission rates, but represent a subset of those admissions. However, they do extend to the entire episode time window. As such, the PAC measures, as submitted, don't create added burden of reporting because the readmissions reported are simply a part of the broader 30-day all-cause readmission measures already endorsed by NQF. Because PAC measures are comprehensive, they include patient safety events as well as other adverse events, including hospitalizations and ED visits during the entire continuum of care. As a result, they are a comprehensive measure of avoidable complications for a specific medical episode. The data collection for all of the HCI3 measures is automated by a software package and is fully harmonized with all other PAC measures. A single download automates creation of all reports related to each of the PAC measures.

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.)
PAC measures are composite measures representing "all-cause harms". They look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence. PACs look at readmissions, emergency room visits, adverse events due to errors of omission or commission. They look at complications that are due to patient safety failures, and also those directly related to the index condition. These are all a cause of significant waste and quality concerns. As such, the measure can provide clinicians with an overall and comprehensive view, in one measure, of all potentially avoidable complications for a patient and drive quality improvement efforts.

For clinicians and facilities increasingly engaged in value-based payment efforts and/or driving quality improvement for population health, the value of a PAC measure over a series of related, but more discrete measures, is that one can better determine if the sources of complications primarily stem from activities within the facility or outside the facility, and the specific nature of the complications that have a higher frequency of occurrence. While individual components of the PAC measure may have small frequencies and may be difficult to interpret with regards to provider performance or actionability, aggregating all the PACs into a comprehensive, composite measure provides the parsimony that is so desirable. For providers, it's far easier to construct a quality dashboard from a parsimonious set of measures, and that's what PAC measures offer.

Further, as a comprehensive outcome measure, PACs are also useful for public transparency of quality, as substantiated by the research from Judy Hibbard and colleagues previously cited in the "testing" section of this submission. As a comprehensive outcome measure, they are easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PACs_and_Severity_Adjustment_Fact_Sheet_HCI3-635719661304570034.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.2 **Point of Contact:** Francois, de Brantes, Francois.debrantes@hci3.org, 203-270-2906-
Co.3 **Measure Developer if different from Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.4 **Point of Contact:** Amita, Rastogi, amita.rastogi@hci3.org, 213-934-9624-

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.
From 2006 onwards, and under the auspices of various funding organizations, HCI3 has convened and managed, or helped to convene and manage, Clinical Working Groups to inform the development and refinement of the measures. For example, in 2011,

2012 and 2013, HCI3 worked collaboratively with the American Board of Medical Specialties and the American Medical Association's Physicians Consortium for Performance Improvement, under a federal contract, to convene and get input from various clinical experts on definitions of episodes of care and their sequelae, including avoidable complications.

Some of the clinical experts that have contributed to our work include:
-Dr. John Allen, American Gastroenterology Association (AGA)
-Dr. Morton Arnsdorf, Cardiologist, University of Chicago, IL
-Dr. Peter Bach, Memorial Sloan Kettering Cancer Center (MSKCC)
-Dr. Peter Basch, Primary Care, Medstar Health, DC
-Dr. Justin Beckelman, Radiation Oncology, University of Pennsylvania, PA
-Dr. Debra Bingham, Executive Director, California Maternal Quality Care Collaborative (CMQCC) at Stanford University, CA
-Dr. John Birkmeyer, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Linda Bosserman, Wilshire Oncology Medical Group, CA
-Dr. Matthew Brengman, American Society of Metabolic and Bariatric Surgery (ASBMS)
-Dr. Joel Brill, American Gastroenterology Association (AGA)
-Dr. George Cautilli, Cautilli Orthopedic Surgical Specialists PC, Yardley, PA
-Dr. Ashwini Davison, Internist, Johns Hopkins Hospital, MD
-Dr. James Denneny, III, American Academy of Otolaryngology – Head and Neck Surgery (AAO-HNS)
-Dr. Chris Gallagher, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Robert Haralson, III, American Academy of Orthopedic Surgeons (AAOS)
-Ms. Dawn Holcombe, Executive Director, Connecticut Oncology Association, CT
-Dr. Colin Howden, American Gastroenterology Association (AGA)
-Dr. John Knightly, American Association of Neurological Surgeons (AANS)
-Dr. Larry Kosinski, American Gastroenterology Association (AGA)
-Dr. Nalini Krishnan, Obstetrics & Gynecology, MN
-Dr. Kelly Kyanko, Internist, NYU School of Medicine, NY
-Dr. Tara Lagu, Internist & Infectious Disease, Baystate Medical Center, MA
-Dr. Robert Lee, Society of Thoracic Surgeons (STS)
-Dr. Alex Little, Society of Thoracic Surgeons (STS)
-Dr. Michael London, Orthopedic Surgeon, OMNI Orthopedics, OH
-Dr. Elliott Main, Obstetrics & Gynecology, California Pacific Medical Center, CA
-Dr. Constantine Mantz, 21st Century Oncology, FL
-Dr. Joseph Messer, Cardiologist, Rush University Medical Center, IL
-Dr. David Metz, American Gastroenterology Association (AGA)
-Dr. Ronald Nahass, Infectious Disease Care, NJ
-Dr. Ajay Nehra, Urologist, Rush University Medical Center, IL
-Dr. Francis Nichols, Society of Thoracic Surgeons (STS)
-Dr. Patrick O'Connor, Primary Care, HealthPartners, MN
-Dr. Sara Perkel, National Comprehensive Cancer Network, PA
-Dr. David Peura, American Gastroenterology Association (AGA)
-Dr. John Ratliff, American Association of Neurological Surgeons (AANS)
-Dr. Steven Schutzer, Connecticut Joint Replacement Institute, CT
-Dr. Leif Solberg, Primary Care, HealthPartners, MN
-Dr. Scott Sporer, Midwest Orthopedics at Rush, Chicago IL
-Dr. Bonnie Weiner, Cardiologist, Worcester Medical Center, MA
-Dr. Jonathan Weiner, Bariatric Surgery codes, Prof of Health Policy and Management, Johns Hopkins University, MD
-Dr. Janet Wright, Cardiologist, Northstate Cardiology Consultants, CA

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** Yearly
Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

#2749 Proportion of Patients with Arrhythmias (ARR) that have a Potentially Avoidable Complication (during the episode time window), Last Updated: Aug 18, 2015

NATIONAL QUALITY FORUM

Measure Information - Composite

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 subcriterion 1b).

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

Brief Measure Information

NQF #: 2749
De.2. **Measure Title:** Proportion of Patients with Arrhythmias (ARR) that have a Potentially Avoidable Complication (during the episode time window)
Co.1.1. **Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)
De.3. **Brief Description of Measure:** Percent of adult population aged 18 + years who triggered an episode of arrhythmias (ARR), are followed for at least one-year, and have one or more potentially avoidable complications (PACs). PACs may occur any time during the episode time window. Please reference attached document labeled NQF_ARR_all_codes_risk_adjustment_06.30.15.xls, in the tabs labeled PACs I-9 and PAC I-10 for a list of code definitions of PACs relevant to ARR.

We define PACs as one of two types:
(1) Type 1 PACs - PACs directly related to the index condition: Patients are considered to have a PAC, if they receive services during the episode time window for any of the complications directly related to ARR, such as for hypotension, cardiac arrest, fluid and electrolyte disturbances etc.
(2) Type 2 PACs - PACs suggesting Patient Safety Failures: Patients are also considered to have a PAC, if they receive services during the episode time window for any of the complications related to patient safety failures such as for sepsis, infections, phlebitis, deep vein thrombosis, pressure sores etc.
All relevant admissions in a patient with ARR are considered potentially avoidable and flagged as PACs.

PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook labeled NQF_ARR_all_codes_risk_adjustment_06.30.15.xls serves as an example. The tab labeled PAC overview gives the percent of ARR episodes that have a PAC and the tab labeled "PAC drill down" gives the types of PACs and their frequencies in ARR episodes within this dataset.

The information is based on a two-year claims database from a large regional commercial insurer. The database had over 3.2 million covered lives and over $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.
1b.1. **Developer Rationale:** Each individual PAC, when measured in isolation, provides a very limited picture of the performance of the provider(s) who are managing or co-managing the care of the patient. However, looking at all the PACs that may occur individually or concurrently in a patient with a given episode provides a comprehensive picture of the care received by the patient for that particular condition or illness.

Additionally, the frequency of occurrence of individual PACs may be so low that it may require very high sample sizes from individual providers to achieve any meaningful and reliable comparisons. But aggregating all the PACs into a single quality metric creates meaningful scores that can be compared across providers even with relatively smaller sample sizes.

Additionally, a comprehensive measure is easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

S.4. **Numerator Statement:** Outcome: Number of patients who triggered an episode of arrhythmias (ARR), are followed for at least one-year, and had one or more potentially avoidable complications (PACs) during the episode time window.
S.7. **Denominator Statement:** Adult patients aged 18 years and above who triggered an episode of arrhythmias (ARR) and are followed for at least one-year.
S.10. **Denominator Exclusions:** Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the ARR measure if they are considered not relevant to ARR care.

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

Is this a MAINTENANCE measure submission? ☐ **Yes** ☒ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: n/a **Most Recent Endorsement Date:** n/a

1d.1. **Composite Measure Construction:** any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)
Component Measures (if endorsed or submitted for endorsement):

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, an
d identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for this *health outcomes* measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if health outcomes measures agree the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

- This new risk-adjusted (by age, gender and clinical co-morbidities) outcomes measure that assesses the proportion of adult patients with claims triggered cardiac arrhythmias (ARR) with at least one Potentially Avoidable Complications (PAC) within 12 months of ARR triggered claims data. Based on NQF's criteria, this measure is considered an "any or none" composite measure that assesses if 1 or more PACs or "care defects" have occurred for the index episode. For this composite measure, the individual complications considered the measurable components
- PACs are classified in two types: 1) related to ARR, and 2) related to Patient Safety Failures, combining the 2 types into a single PAC rate to calculate the proportion of patient with 1 or more PAC. PACs are considered unwarranted health outcomes that combine concepts from AHRQ PSIs, PQIs and the CMS HACs and episode-specific PACs into all-cause patient harms that is measured during an index condition for use at the practice, medical group, provider system or purchaser/payer levels to identify quality of care gaps between practices and hospitals.
- The developer links primary & secondary prevention care gaps, poor patient education, poor care coordination and poor follow-up increase unnecessary ER visits, hospitalizations, readmissions, and mortalities to increased PACs, and

state that PACs for ARR patients should occur rarely in well-managed patients.

- The evidence for Patient Safety Failure PACs is described to be within the influence of the measured entity, and does not describe the influence of non-healthcare-related impacts on PAC rates. The progression of the episode condition, illness or disease is also not mentioned as a contributor to PAC rates in the evidence.
- In addition to linking processes of care to outcomes, the developer provides an extensive PAC literature review in sections 1a.2. and 1a.2.1. for ARR, Patient Safety Failures & processes of care, as well as background information on the process for PAC development.

Questions for the Committee:

- *Does sufficient evidence exist connecting Patient Safety Failures to the ARR index episode?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*

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

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

- The developer provides ARR prevalence & impacts data, rationales and general information on PAC measure utility and applicable setting use. The developer identifies ARR as a predictor of stroke, and provides 2012 stoke & ARR prevalence, prevention, cost and readmission data.
- ARR PAC performance gap data are calculated from PROMETHEUS administrative claims data from April 1, 2012 through December 17, 2014, for providers with ≥ 10 attributable index episodes. The data includes 575 of 6,728 (8.5%) providers from 38,207 of 81,216 (47.0%) index episodes in 3,258,706 unique beneficiaries.

Unadjusted PAC Rates:		**Risk-Standardized PAC Rates (RSPR):**	
Median (IQR):	35.7% (26.3%, 45.5%)	Median (IQR):	35.9% (29.2%, 42.1%)
Range:	0% - 100%	Range:	0% - 91.9%

- Limited descriptive data on the patient, provider and payer are provided. The developer provides "Overview" and "Drill Down" PAC rates in the spreadsheet demonstrating gap, though PAC rates for individual complications are not provided.
- No disparities data was provided.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - None

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- The evidence to support the measure shows from prior studies that gaps in dysrhytmia management can lead to potentially avoidable complications.

1b. ***Committee's Comments on Performance Gap:***

- The developer provides ARR prevalence and impact data, rationale and general info on PAC measure utility.
- PAC date calculated from the PROMETHEUS admin claims data showed an unadjusted PAC rate vs. risk standardized PAC rate of 35.7% vs. 35.9%, however no disparity data was provided.

1c. ***Committee's Comments on Composite Performance Measure:***

- They are not clearly stated and logical.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure assesses the rate of patients with 1 or more PAC(s) during index episodes. This new risk adjusted outcomes measure is specified for use at the clinician group and team levels of analyses.
- The measure exclusively uses electronically available administrative claims data to calculation the measure score, and for this measure, better care equals lower scores.
- The developer describes non-patient-related PACs as controllable by provider processes without further analysis as other influencers that may contribute to PAC rates beyond the patient and provider (e.g., payer, access, suppliers, etc.).
- Patient- and claims-based exclusions are provided to promote the availability and consistency of claims data capture, including payer enrollment requirements, cost outliers of < 1% or > 99%, and claims not relevant to ARR.
- Developers provide administrative claims for ARR & PAC (ARR - & Patient Safety Failure-related) triggers, describe a complete 12-month episode time window. A calculation algorithm.
- ICD-9 & ICD-10 codes are provided, though ICD-10 descriptions & an ICD-9 to ICD-10 crosswalk methodology are not provided.
- A conceptual risk model and statistical method using logistic regression model for determining the probability of a patient incurring a PAC are provided. After adjusting for patient-related factors, the developers state the remaining PAC variance is due to factors potentially controlled by the provider during and after hospitalization. "Predicted" coefficients from the risk adjustment models are summed to give predicted probabilities of PAC occurrence.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is this measure specified to pertain only to providers with at least 10 episodes (per the reliability testing described below)?*
- *Is it likely this measure can be consistently implemented?*
- *Is additional evidence required to determine whether group/practice/team level of analysis is appropriate?*

2a2. Reliability Testing Testing attachment

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.

- The developer tested reliability at the performance measure score, and used a beta-binomial model and a signal-to-noise analysis, which is appropriate for this type of measure, to differentiate the true difference between measured entities (the signal) to random measurement error (the noise). A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in between hospital performance. A value of 0.7 is often regarded as a minimum acceptable reliability value.
- The measure is specified for patients with arrhythmias ≥ 18 years, though the testing sample includes patients 18 through 64 years.
- Providers with < 10 ARR episodes were excluded from reliability testing, though the measure is specified for patient without episode restrictions A sample of 5840 providers was initially included in the data set, though providers with less than 10 ARR episodes were excluded, allowing for 468 remaining providers. There were 38,207 episodes of ARR with a mean age of 49.1 (18-64 years) and 54% female in the testing analysis exclusively using administrative claims data.
- The developer states, "Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset will apply to another." The developer also states that very high sample sizes are to achieve any meaningful and reliable comparisons.
- A patient may have more than one condition-specific concurrent episode with claims applied to both episodes. If an inpatient claim corresponds to a procedure index episode and to a condition index episode, the claim would be assigned to the procedure index episode, rather than the condition index episode (e.g., for a claim that corresponds to both index episodes of HF & CABG, the claim would be assigned to CABG).
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- Reliability results are provided in the table below, as well as in great detail in the accompanied spreadsheet with median (IQR) results demonstrating reliability of 0.66 (0.54,0.79) for ≥ 10 providers, increasing with the number of providers, demonstrating the measure is able to demonstrate differences in performance. For reliability analysis, providers were restricted to the minimum of 10 ARR episodes, though all episodes were included in the risk model.

Reliability Scores	Minimum # Episodes Per Provider		
	>=10	>=25	>=50
# of Providers (%)	575 (100)	232 (40)	103 (18)
Median (IQR)	0.66 (0.54,0.79)	0.80 (0.74,0.87)	0.88 (0.84,0.93)
Range	0.42-1.00	0.65-0.99	0.79-0.99

The table provides a summary reliability scores minimum sample size thresholds. Complete results are provided in the workbook entitled, NQF_ARR_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab.

Questions for the Committee:

- *Reliability testing was conducted only for those providers with at least 10 episodes. Can differences in performance be identified for providers with fewer than 10 episodes? Should the measure be specified to include only those providers with at least 10 episodes? Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- Because this is an outcome measure, the rationale that is presented for subcriterion 1a does not necessarily have to address all of the variables used to calculate the measure.
- The measure uses a statistical risk model with 170 risk factors and episode-specific subtypes/severity markers including age, gender and clinical comorbidities, on at least 10 claims to determine "stable" covariates and assess comorbidity or procedure impact on the PAC. All covariates must be present prior to an episode trigger. No formal covariate analysis was conducted to select risk factors beyond the present on a minimum of 10 claims threshold. The developer describes the heterogeneity of the provided data sets as crucial to ensure measure validity, and the accuracy and completeness of the data sets.
- The developers did not provide disparities data, an exploration of a conceptual relation to SDS, or SD S factors in the risk model.

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Are these variables available and generally accessible for the measured patient population?*

2b2. Validity testing

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.

- The developer conducted systematic assessment of face validity for the performance measure score for validity testing in numerous ways, including the use of monthly multi-specialty clinical working groups, and other tests of face validity, along with focus groups, face validity comparisons of the measure to other national accountability measures, as well as additional literature for the measure & PAC development process.
- No empiric results are provided for the face validity tests described above.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
-

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer describes patient- (demographic, enrollment or low/high claims cost) and claims-based (due to missing or non-relevant data) exclusions for the measure. They further state nearly half of the original population of ARR patients was removed from the denominator with applied exclusions.
- A significant number of patients were eliminated from the measure due to exclusion criteria, including 38,207 of 81,216 (47.0%) ARR episodes (in 3,258,706 unique beneficiaries) and 575 of 6,728 (8.6%) providers.

Questions for the Committee:

- *Are high cost outliers (> 99%) exclusions an opportunity to identify PACs?*
- *Does the high number of exclusions restrict the measure use?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*

- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- The risk model (detailed in the accompanied spreadsheet) includes 170 factors and subtypes including age, gender, 12-month enrollment markers, co-morbidities, and episode severity markers.
- No SDS factors beyond age and gender was included in the risk-adjustment approach. Beyond noting that race was not available for analysis, no description of the of the conceptual relationships between patient sociodemographic factors, patient clinical factors, quality of care, and the outcomes (PAC rates) were provided, nor do they discuss a conceptual relationship or variables of SDS to the risk model.
- Logistic regression was used to model the probability of at least one PAC during an episode. The reasoning for no additional modeling performed is described.
- The performance of the model was determined with a split sample method by estimating the model coefficients using a development dataset (80% of the sample) and applying those coefficients to a validation dataset (20% of the sample). C-statistics for the development and validation samples with c-statistic results of 0.781 and 0.773, respectively. C-statistics measures the extent of a statistical model to discriminate between a patient with and without PAC, with an ability to predict if a PAC is or is not present about 75% of the time. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong.
- Both Hosmer-Lemeshow Goodness-of-Fit statistics and risk-decide plots were provided to indicate model fit. Results from the Hosmer-Lemeshow test suggest that the fit is not good; however, this test is sensitive to the number of groupings and sample sizes. Results from the risk decile plot indicate that the predicted PAC rates are similar to the observed PAC rates across all deciles of risk. The developer states the model demonstrates strong predictive power.

Questions for the Committee:

- *Is the Committee aware of conceptual relationship(s) between additional patient-level SDS factors, patient clinical factors, quality of care, and PACs (other than gender and age)? If so, what data might be available to allow an empirical analysis of these relationships?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*

2b5. Meaningful difference:

- The developer presents PAC rates across providers and also providers adjusting for differences in patient severity in a ratio of observed to expected attributable episodes to PACS accounting for patient severity, and calculates estimates from the risk model, for risk-standardized PAC rates for the provider.

Summary of Unadjusted and Adjusted Performance Scores Across Providers

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25
Unadjusted		
Median (IQR)	36% (26%, 45%)	34% (28%, 44%)
Range	0%-100%	6%-98%
Adjusted (RSPR)*		

Median (IQR)	36% (29%, 42%)	36% (31%, 41%)
Range	0%-92%	8%-67%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_ARR_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- As there is only one data source used for measure calculation (administrative claims), comparability of data sources or methods is not applicable.

Question for the Committee:

- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*

2b7. Missing Data

- No formal analysis of missing data is provided. As the measure assesses the rate of patients with PACs, rather than the rate of PACs per index episode, the total number of PACs is not included in the PAC rate.
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- The developers state the under-coding of claims is unavoidable in the current DRG payment structure which could lead to under capture or missing PACs.

2d. Empirical Analysis to Support Composite Construction

- As an "any or none" composite, the individual complications are considered measurable components of the composite. Frequency and distribution statistics are provided in the PAC Overview and PAC Drill Down tabs in the measure spreadsheet, which detail PAC types and subtypes. The identification of individual PACs are not provided (e.g., sepsis, unattended falls, DVT).
- PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.
- The developer states that no formal analysis was performed on missing data. For details, see 2b7 above.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- The conceptual risk model which is a logistic regression model is not clearly defined.

2a2.: ***Committee's Comments on Reliability-Testing:***

- The reliability testing performed had a value of less than 0.7

2b1.: ***Committee's Comments on Validity-Specifications***:

- The risk model using 170 risk factors, it does not show if all of these risk factors are available for every patient.

2b2.: ***Committee's Comments on Validity-Testing:***

- Validity was not tested well enough to allow for generalization

2b3-7.: ***Committee's Comments on Threats to Validity:***

- High number of exclusions will affect the data
- Once again the risk model will not be effective due to possible missing data.

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than the person obtaining the original information (e.g., DRG, ICD-9 codes on claims).
- The developer provides an excel spreadsheet attachment including diagnoses, visits, hospitalizations, post-acute facility stays, procedures, laboratory tests and procedures/surgeries, for ARR & PAC triggers, and describe the time window for measuring PACs as 12 months following a ARR episode triggers, as well as a decision tree for measure calculation and implementation.
- The measure specifications, metadata and calculation algorithms are available free of charge on the developer's website. Limited analytics are planned at no cost to the end user.
- This is not an eMeasure.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

Committee's Comments on Feasibility:

- Data elements are not routinely generated and used during care delivery.

Criterion 4: Usability and Use

4. Usability and 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.

- This is a newly developed claims measure is current used in accountability programs for payers, states, and planned for public reporting, professional certification or recognition programs, and external quality improvement for benchmarking purposes.
- The developer states that PAC measures provide a foundation for the relationship between healthcare quality and cost and assist in the exploration of practice reengineering and alternative payment models, act as indicators of potential harm, and is spurring the development of private-based analytics software for further outcomes exploration. No public improvement rates are available due to recent implementation and variation in PAC definitions have also modified, though the provided PROMETHEUS data suggest wide variation in performance and improvement opportunities. Payer and Provider improvement use perspectives are also outlined.
- The developer found no noted unintended consequences, though they state the measure is intended for transparency

and QI activities only. They also state the under-coding of claims is unavoidable in the current DRG payment structure could be an unintended consequences of the measure, and payers calculating the measures even with inadequate sample sizes and using the results to penalize providers could lead to invalid provider comparisons.

- If the measure was theoretically to be used for accountability purposes to "ding" the measured entity as defined in the level of analysis, further exploration of PAC antecedents and the measured entity is warranted, especially with small group practices and very small PAC rates. In the original testing sample of 6,728 providers, when providers with fewer than 10 ARR episodes were eliminated from analysis due to less reliability estimates with small numbers, 575 (8.6%) remained for analysis.

***Questions for the Committee**:*

- *Is the measure publicly reported?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Should PAC measures also include the clinician: group in the analysis or include population-level only entities?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- It may be difficult to implement the measure as it is proposed

Criterion 5: Related and Competing Measures

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

0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2749

Measure Title: Proportion of Patients with Arrhythmias (ARR) that have a Potentially Avoidable Complication (during the episode time window)

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: 6/30/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading

definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☒ Health outcome: Potentially Avoidable Complications

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

There are significant gaps in the management of patients with dysrhythmias leading to potentially avoidable complications (PACs) contributing to unnecessary inefficiencies and waste in health care with the consequent increased socioeconomic burden. The Framingham study showed a 23.5% risk of stroke attributable to AF (Wolf 1991). Despite this a shocking 30% to 50% of eligible AF patients do not receive preventive anticoagulation (Reynolds 2012). The use of optimal anticoagulation therapy in just half of these patients could prevent 19,000 strokes and save more than $1.1 billion in direct costs annually. Additionally, frequent hospitalizations are common in patients with AF and a study showed that 1 in 8 patients are readmitted (Kim 2005).

There is enough evidence in the literature that highlights significant "gaps in care" in management of patients with chronic conditions (McGlynn 2003). Gaps in care, in turn lead to process failures that cause patients to incur unnecessary services and some harm (Jha 2013). For example, a report by the Agency of Health Care Research and Quality (AHRQ) highlighted the fact that in 2008, $4.4 million out of a total of 39 million (11 percent) hospital-stays that could have been prevented (Stranges 2008); and for Medicare beneficiaries one in five admissions were for a potentially preventable condition (Jiang 2006). To improve accountability in the delivery of medical care, AHRQ has developed a list of patient safety indicators (PSIs) to identify potential harms to patients (Miller 2001). Additionally, the Centers for Medicare and Medicaid Services (CMS) have

taken a "Six Sigma" approach and defined Hospital Acquired Conditions (HACs) and "never events" that should almost never occur and are applying financial penalties when these events do occur (CMS 2012).

The Potentially avoidable complications (PAC) measure goes beyond the AHRQ PSIs and the CMS HACs and creates a single comprehensive measure that measures all-cause harms for a patient with the index condition. Potentially avoidable complications (PACs) are the unwarranted health outcomes that this measure addresses (deBrantes 2010). Lack of patient education on self care techniques, poor care coordination, and poor arrangements of patient follow-up could lead to unnecessary ER visits, hospitalizations and gaps in care leading to increased morbidity. All these adverse events are aggregated together as a single comprehensive measure to study the overall rate of PACs in the arrhythmia / heart block population.

Adult patient diagnosed with Arrhythmia or a heart block (ARR)

↓

Physician practices fail to educate patients / Physician practices have poor access

↓

Patient visits ER / gets hospitalized

↓

Patient discharged with management advise / remains in hospital for treatment of PAC

Well-managed patients with ARR should rarely incur a potentially avoidable complication such as an emergency room visit, and hospitalizations related to ARR should occur only in the rarest of circumstances.

The enclosed workbook entitled NQF_ARR_all_codes_risk_adjustment_06.30.15.xls lists the types of PACs and their frequency as calculated in a large regional database (see tab PAC overview). Over 36% of patients with ARR had a PAC, with about 12% of PACs directly related to ARR itself, such as fluid and electrolyte disorders, acute heart failure or pulmonary edema etc.(see tab PAC Drill Down Graph). Although the preventable hospitalizations in the ARR population were low, at only 8.2% of all ARR episodes; approximately 30% of patients with ARR had PACs related to patient centered care failures such as poor control of diabetes, respiratory insufficiency and acute gastritis, many of them being managed in an outpatient setting in physician offices. As a result over 34% of episodes had a PAC indicator on the professional claims.

While PACs may not be eliminated completely, identifying the magnitude of PACs and knowledge of the cause for the most frequent or the most expensive PACs could place an emphasis in reducing them and as a consequence improving patient outcomes. The ability to clearly identify the type and frequency of each PAC, creates a highly actionable measure for all providers that are managing or co-managing the patient; as well as for the health plan with whom the patient is a member (de Brantes 2009).

References:

1) Wolf PA, Abbott RD, Kannel WB. "Atrial fibrillation as an independent risk factor for stroke: the Framingham Study." *Stroke* 22.8 (1991): 983-988.

2) Reynolds, Matthew R. MD, MSc; and Vidal Essebag, MD, PhD "Economic Burden of Atrial Fibrillation: Implications for Intervention" *Am J Pharm Benefits* 4.2 (2012):58-65. Web.

3) Kim MH, Lin J, Hussein M, Battleman D. "Incidence and temporal pattern of hospital readmissions for patients with atrial fibrillation." *Curr Med Res Opin* 25.5 (2009):1215-1220.

4) McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348.26 (2003): 2635–45.

5) Jha, A. K., I. Larizgoitia, C. Audera-Lopez, N. Prasopa-Plaizier, H. Waters, and D. W. Bates. "The Global Burden of Unsafe Medical Care: Analytic Modeling of Observational Studies." *BMJ Quality & Safety* 22.10 (2013): 809-15. Web.

6) Stranges, Elizabeth, MS, and Carol Stocks, RN, MHSA. "Potentially Preventable Hospitalizations for Acute and Chronic Conditions, 2008." *Agency for Healthcare Research and Quality* 99 (2010). Web.

7) Jiang HJ, Russo CA, and Barrett ML. Nationwide Frequency and Costs of Potentially Preventable Hospitalization, 2006. AHRQ-HCUP Statistical Brief # 72. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb72.jsp.

8) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." *Heath Services Research* 36.6.2 (2001): 110-132.

9) Centers for Medicare and Medicaid Services. "Hospital Acquired Conditions." *CMS.gov*. Centers for Medicare and Medicaid Services, 2012. Web.

10) de Brantes, Francois, Amita Rastogi, and Michael Painter. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." *Health Services Research 2nd ser.* 45.6 (2010): 1854-871. Web.

11) de Brantes, Francois M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. "Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model." *NEJM* (2009) 361:1033 (Perspective)

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

Rationale: Better processes of care create an atmosphere of proactive management, consistency in care and standardized care patterns (Shekelle 2013) (Fenter 2006). Patient education and adopting safe practices significantly reduces occurrence of potentially avoidable complications (PACs) in all settings (Klein 2011) (Wachter 2013) (Berwick 2006) (Kovner 2011) (Farley 2013). It is known that by holding providers accountable for occurrence and costs of PACs, an built-in warranty is created around care of the index condition (de Brantes 2009).

In 2014, the AHA/ACC/HRS published their evidence-based guidelines for management of patients with Atrial Fibrillation with an aim to prevent the consequences of poorly managed arrhythmia, namely, stroke, congestive heart failure, increased hospitalizations, cognitive decline and death (Stewart 2002) (Wolf 1991) (Benjamin 1998). Additionally, the 2012 ACCF/AHA/HRS Focused Update of the 2008 Guidelines for Device-Based Therapy (DBT) of Cardiac Rhythm Abnormalities is a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines for device use in patients with arrhythmias and heart block (Tracy 2012). Utilization of these evidence-based guidelines has a potential to reduce unnecessary hospitalizations and complications in patients with arrhythmias. The use of optimal anticoagulation therapy could by itself lead to prevention of some of the major complications of heart rhythm abnormalities. Advances in understanding the mechanisms underlying rhythm abnormalities, clinical implementation of ablation techniques to maintain sinus rhythm and newer drugs for stroke prevention, if incorporated appropriately into clinical practice, could potentially bring down the rate of complications.

Data from the Healthcare Cost and Utilization Project indicate that roughly 60% of hospital admissions for a principal diagnosis of AF result from emergency department visits (HCUP). Evidence suggests that AF can be safely treated on an outpatient basis and unnecessary hospitalizations can be avoided. For example, electrical cardioversion can be safely performed as an outpatient at a much lower cost than as an inpatient (Birger 2003).

Hospitalizations in patients for medical management of Arrhythmias or heart blocks should be rare and mostly avoidable. However, if patients do get hospitalized, careful management while in the hospitals and good discharge planning and follow-up prevents unnecessary complications, ER visits and readmissions (Weaver 2013) (Mittler 2013). Another study from the Boston Medical Center, demonstrated that although one in five hospitalizations are complicated by post-discharge adverse events, development of a strong discharge services program for patients admitted for medical conditions reduced hospital utilization within 30 days of discharge (Jack 2009). In addition, while in the hospital, safe practices reduce the burden of healthcare associated complications (Ranji 2007). Some of these are listed below:

1. Frequent hand-washing reduce hospital acquired infections (WHO 2007)
2. Carefully implemented protocols lead to reduced line sepsis (Pronovost 2010)
3. Discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013)
4. DVT prophylaxis in patients on bed rest avoids pulmonary embolism (Shekelle 2013)
5. Frequent change in position of ARR patients in the CCU avoids pressure sores (Shekelle 2013)

PAC measures in the setting of arrhythmias and / or heart blocks look at all-cause harms, such as the ones highlighted above, arising from poor management of these patients.

References:

1) Shekelle, Paul G., et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

2) Fenter TC, et al. "The cost of treating the 10 most prevalent diseases in men 50 years of age or older." *Am J Manag Care* 12.4 (2006): S90-S98.

3) Klein S., and D. McCarthy, OSF HealthCare: "Promoting Patient Safety Through Education and Staff Engagement." *The Commonwealth Fund* 1475.7 (March 2011).

4) Wachter, Robert M., Peter Pronovost, and Paul Shekelle. "Strategies to Improve Patient Safety: The Evidence Base Matures." *Annals of Internal Medicine* 158.5_Part_1 (2013): 350-52. Web.

5) Berwick, D.M., D. R. Calkins, C. J. McCannon et al., "The 100,000 Lives Campaign: Setting a Goal and a Deadline for Improving Health Care Quality," *Journal of the American Medical Association*,295.3 (2006):324–27

6) Kovner, Anthony R., James Knickman, and Steven Jonas. "Chapter 11." *Jonas & Kovner's Health Care Delivery in the United States*. 10th ed. New York: Springer Pub., 2011. Print.

7) Farley, TG, et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

8) de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? *Health Affairs (Millwood)* 28 (2009): w678-w687.

9) Stewart S, Hart CL, Hole DJ, McMurray JJ. "A population-based study of the long-term risks associated with atrial fibrillation: 20-year follow-up of the Renfrew/Paisley study." *Am J Med* 113.5 (2002): 359-364. Web.

10) Wolf PA, Abbott RD, Kannel WB. "Atrial fibrillation as an independent risk factor for stroke: the Framingham Study." *Stroke* 22.8 (1991): 983-988.

11) Benjamin EJ, et al. "Impact of atrial fibrillation on the risk of death: the Framingham Heart Study." *Circulation* 98.10 (1998): 946-952.

12) Tracy Cynthia M., et al. "The 2012 ACCF/AHA/HRS Focused Update of the 2008 Guidelines for Device-Based Therapy of Cardiac Rhythm Abnormalities, A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines." *J Am Coll Cardiol* 60.14 (2012):1297-1313.

13) Birger, Sally, Botkin, Leela S. Dhanekula, Brian Olshansky. "Outpatient Cardioversion of Atrial Arrhythmias: Efficacy, Safety, and Costs" *Am Heart J* 145.2 (2003) Web.

14) Weaver, Sallie J., Lisa H. Lubomksi, Renee F. Wilson, Elizabeth R. Pfoh, Kathryn A. Martinez, and Sydney M. Dy. "Promoting a Culture of Safety as a Patient Safety Strategy." *Annals of Internal Medicine* 158.5_Part_2 (2013): 369-75. Web.

15) Mittler, Jessica N., Jennifer L. O'hora, Jillian B. Harvey, Matthew J. Press, Kevin G. Volpp, and Dennis P. Scanlon. "Turning Readmission Reduction Policies into Results: Some Lessons from a Multistate Initiative to Reduce Readmissions." *Population Health Management 16*.4 (2013): 255-60. Web.

16) Jack BW, Chetty VK, Anthony D et al. "A Reengineered Hospital Discharge Program to decrease Rehospitalization: A randomized trial." *Ann Int Med* 150 (2009):178-187.

17) Ranji SR, Shetty K, Posley KA, et al. "Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol 6 Prevention of Healthcare-Associated Infections) Technical Review 9" *AHRQ* 04(07)-0051-6. January 2007.

18) The Joint Commission, Aide Joint Commission International, and WHO. "Improved Hand Hygiene to Prevent Health Care-Associated Infections." World Health Organization. WHO Collaborating Centre for Patient Safety Solutions (May 2007). Web.

19) Pronovost, P.J., G. A. Goeschel, E. Colantuoni et al., "Sustaining Reductions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study," *BMJ* (2010): 340:c309 Web.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

2749_ARR_Evidence_Attachment_HCI3-635717850211232805.docx

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., the benefits or improvements in quality envisioned by use of this measure)*

Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013 (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 50% of its plan members with arrhythmia / heart block incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow-up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative

data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) de Brantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.

4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

6) BCBSNC: Blue Cross Blue Shield of North Carolina: https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5. http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/

11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.

14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR.

Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The data included two years of administrative claims covering the period April 1, 2012 through December 17, 2014. There were a total 38,207 episodes of ARRBLK.

Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. After this exclusion 575 (out of 6728) providers remained. Performance scores of these providers are summarized in the following table:

Unadjusted PAC Rates:
Median (IQR): 35.7% (26.3%, 45.5%)
Range: 0% - 100%

Risk-Standardized PAC Rates (RSPR):
Median (IQR): 35.9% (29.2%, 42.1%)
Range: 0% - 91.9%

Please refer to the NQF_ARRBLK_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

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.
The economic burden to Medicare for stroke in AF patients is estimated to be $2.6 billion. Atrial fibrillation (AF) is an independent predictor of stroke and heart failure. Stroke, in particular, is more severe and costly in AF than in non-AF patients. The Framingham study showed a 23.5% risk of stroke attributable to AF (Wolf 1991). Despite this a shocking 30% to 50% of eligible AF patients do not receive preventive anticoagulation (Reynolds 2012). The use of optimal anticoagulation therapy in just half of these patients could prevent 19,000 strokes and save more than $1.1 billion in direct costs annually.

A retrospective analysis of 3 federally funded US databases showed that cardiac dysrhythmias particularly AF majorly contributes to the substantial health care cost burden of $6.65 billion annually (2005 US dollars) and up to 75% of this cost is attributable to the direct and indirect costs of hospitalizations (Coyne 2006). Frequent readmissions are common in patients of AF and a study showed that 1 in 8 patients are readmitted (Kim 2009). In a retrospective analysis by Coyne et al., the incremental costs due to AF per hospitalization for congestive heart failure and acute myocardial infarction were $1682 and $4422, respectively, and incremental costs due to AF annually for congestive heart failure and acute myocardial infarction were $372,060,082 and $243,917,520, respectively (Coyne 2006).

More broadly, the June 2007 MedPAC report to Congress on "Promoting Greater Efficiency in Medicare" highlighted the fact that in 2005, $12 billion were spent on potentially preventable readmissions alone within 30 days of discharge from the hospital. Another study by Jencks and colleagues found that roughly 19.6% of Medicare patients incurred re-hospitalizations within 30 days of discharge. When hospitalizations do occur, they must be managed expeditiously and readmissions following discharge should be avoided (MedPac 2007).

While PACs may not be completely eliminated, identifying their magnitude and understanding their causality, in particular for the most frequent or the most expensive, could lead to improving patient outcomes.

References:
1) Wolf PA, Abbott RD, Kannel WB. "Atrial fibrillation as an independent risk factor for stroke: the Framingham Study." Stroke 22.8 (1991): 983-988.

2) Reynolds, Matthew R. MD, MSc; and Vidal Essebag, MD, PhD "Economic Burden of Atrial Fibrillation: Implications for Intervention" Am J Pharm Benefits 4.2 (2012):58-65. Web.

3) Coyne KS, et al. "Assessing the direct costs of treating nonvalvular atrial fibrillation in the United States." Value Health 9.5 (2006): 348-356.

4) Kim MH, Lin J, Hussein M, Battleman D. "Incidence and temporal pattern of hospital readmissions for patients with atrial fibrillation." Curr Med Res Opin 25.5 (2009):1215-1220.

5) MedPac. Report to the Congress: Promoting Greater Efficiency in Medicare. Rep. Washington DC: Medicare Payment Advisory Commission, 2007. Web.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Not applicable

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
An estimated 2-3 million Americans suffer from atrial fibrillation (AF) making it the most common cause of cardiac arrhythmia encountered in clinical practice. The incidence of cardiac arrhythmias including heart block increases with age. The incidence of third-degree heart block is highest in people older than 70 years. With the aging population of baby boomers and improved survival of patients with cardiac disease, the prevalence of atrial fibrillation is estimated to increase to 15.9 million cases by the year 2050 and more than half of these patients will be 80 years or older (Reynolds 2012).

There are significant gaps in the management of patients with dysrhythmias leading to potentially avoidable complications (PACs) contributing to unnecessary inefficiencies and waste in health care with the consequent increased socioeconomic burden. The consequences of poorly managed arrhythmias are thromboembolic disease, particularly stroke, 3-3.5 fold increase in risk of heart failure, increased hospitalizations and all cause mortality (Stewart 2002) (Wolf 1991) (Benjamin 1998).

The economic burden to Medicare for stroke in AF patients is estimated to be $2.6 billion. Atrial fibrillation (AF) is an independent predictor of stroke and heart failure. Stroke, in particular, is more severe and costly in AF than in non-AF patients. The Framingham study showed a 23.5% risk of stroke attributable to AF (Wolf 1991). Despite this a shocking 30% to 50% of eligible AF patients do not receive preventive anticoagulation (Reynolds 2012). The use of optimal anticoagulation therapy in just half of these patients could prevent 19,000 strokes and save more than $1.1 billion in direct costs annually.

A retrospective analysis of 3 federally funded US databases showed that cardiac dysrhythmias particularly AF majorly contributes to the substantial health care cost burden of $6.65 billion annually (2005 US dollars) and up to 75% of this cost is attributable to the direct and indirect costs of hospitalizations (Coyne 2006). Frequent readmissions are common in patients with AF and a study showed that 1 in 8 patients are readmitted (Kim 2009). In a retrospective analysis by Coyne et al., the incremental costs due to AF per hospitalization for congestive heart failure and acute myocardial infarction were $1682 and $4422, respectively, and incremental costs due to AF annually for congestive heart failure and acute myocardial infarction were $372,060,082 and $243,917,520, respectively (Coyne 2006).

Therefore, there are many areas where improvement is possible in patients with arrhythmia or heart block, making it a high priority aspect of health care. The PAC measures go beyond simple readmission rates and look for all-cause harms in these patients.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1) Reynolds, Matthew R. MD, MSc; and Vidal Essebag, MD, PhD "Economic Burden of Atrial Fibrillation: Implications for Intervention" Am J Pharm Benefits 4.2 (2012):58-65. Web.

2) Stewart S, Hart CL, Hole DJ, McMurray JJ. "A population-based study of the long-term risks associated with atrial fibrillation: 20-year follow-up of the Renfrew/Paisley study." Am J Med 113.5 (2002): 359-364. Web.

3) Wolf PA, Abbott RD, Kannel WB. "Atrial fibrillation as an independent risk factor for stroke: the Framingham Study." Stroke 22.8 (1991): 983-988.

4) Benjamin EJ, et al. "Impact of atrial fibrillation on the risk of death: the Framingham Heart Study." Circulation 98.10 (1998): 946-952.

5) Coyne KS, et al. "Assessing the direct costs of treating nonvalvular atrial fibrillation in the United States." Value Health 9.5 (2006): 348-356.

6) Kim MH, Lin J, Hussein M, Battleman D. "Incidence and temporal pattern of hospital readmissions for patients with atrial fibrillation." Curr Med Res Opin 25.5 (2009): 1215-1220.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable

1d. Composite Quality Construct and Rationale

1d.1. **A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
  - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient); or
  - any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient).

1d.1. Please identify the composite measure construction: any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

1d.2. **Describe the quality construct, including:**

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

We classify PACs into two types: Type 1 PACs are directly related to the index condition and are often controlled by the servicing provider; Type 2 PACs, on the other hand result from patient safety failures and could be reduced by better systems and better processes in care. Both types of PACs could occur in any setting and so could be identified through any type of claims coming in the administrative dataset, including in-patient, out-patient, or professional claims. PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1.

The PAC measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs), Hospital Inpatient Quality Reporting measures, Avoidable Readmissions, AHRQ defined patient safety indicators (PSIs), NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates.

All defined PACs, irrespective of their type, or site of occurrence, are aggregated to create an overall comprehensive, composite measure. They all have equal weighting, since they are measured simply by the frequency of their occurrence.

1d.3. **Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.**

Each individual PAC, when measured in isolation, provides a very limited picture of the performance of the provider(s) who are managing or co-managing the care of the patient. However, looking at all the PACs that may occur individually or concurrently in a patient with a given episode provides a comprehensive picture of the care received by the patient for that particular condition or illness.

Additionally, the frequency of occurrence of individual PACs may be so low that it may require very high sample sizes from individual providers to achieve any meaningful and reliable comparisons. But aggregating all the PACs into a single quality metric creates meaningful scores that can be compared across providers even with relatively smaller sample sizes.

Additionally, a comprehensive measure is easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

1d.4. **Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**

In constructing the comprehensive composite PAC measure, each component PAC, as clinically defined by the subject matter experts, was given the same weight so that arbitrary weights may not bias the results. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence. As such, the patient is the ultimate unit of measurement and if the patient incurred any PAC during the episode, then that counts against the numerator.

Since the emphasis of the PAC measure was to simply identify the occurrence of PACs in any setting, aggregation of the PAC counts to create a comprehensive quality score with equal weights has been met with overall support from the clinical working groups as well as from the implementation sites.

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 subcriteria 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 : Atrial Fibrillation

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Healthcare Associated Infections, Safety : Medication Safety, Safety : Readmissions, Safety : Venous Thromboembolism

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.)*
http://www.hci3.org/ecr_descriptions/ecr_description.php?version=5.2.006&name=ARRBLK&submit=Submit

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_ARR_all_codes_risk_adjustment_06.30.15-635719684977642819.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Outcome: Number of patients who triggered an episode of arrhythmias (ARR), are followed for at least one-year, and had one or more potentially avoidable complications (PACs) during the episode time window.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The time window is the most recent 12 months of the episode, once a patient has triggered an ARR episode.

S.6. **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, 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.
Patients that have triggered a ARR episode, and are identified as having services for potentially avoidable complications (PACs), during the most recent 12 months of the episode. The enclosed excel workbook entitled NQF_ARR_all_codes_risk_adjustment_06.30.15 gives the detailed codes for PACs in the tabs entitled PACs I-9 and PACs I-10. PACs are identified only based on diagnosis codes.

Services for PACs are identified as follows:
a. Any service (professional, outpatient facility, ancillary) that is relevant to ARR and has a PAC code in any position on the claim
b. Any admission to an acute care facility, that is relevant to ARR
c. Any admission to a post-acute care facility that is relevant to ARR and has a PAC code in any position on the claim

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*

Adult patients aged 18 years and above who triggered an episode of arrhythmias (ARR) and are followed for at least one-year.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Please refer to the enclosed excel workbook entitled
NQF_ARR_all_codes_risk_adjustment 06.30.15

The target population is identified using the following criteria:
1.Using administrative claims database, patients with ARR are identified using one of the following criteria:
a. Patients having an office visit with a trigger code of ARR in any position, followed by a second confirmatory office visit (with a trigger code of ARR in any position), at least 30 days apart.
b. Patients with a Principal Dx of an ARR trigger code on an in-hospital stay claim.
The trigger codes for ARR are provided in the tab called "Triggers I-9" or "Triggers I-10".
2. The patient should have continuous enrollment for the entire time window with no more than 30 days as an enrollment gap, with the entity providing the data (so we can ensure that the database has captured most of the claims for the patient during the episode time window).
3. The patient should have a complete episode time window in the claims data – so there are at least 12 months of claims in the database for the patient.
4. Patient should be at least 18 years of age

Once the episode is triggered all relevant claims are assigned to the episode. Relevant claims could be inpatient facility claims, outpatient facility claims, professional services, laboratory services, imaging services, ancillary claims, home health, durable medical equipment as well as pharmacy claims across the entire continuum of care centered around the patient's episode of care. Relevant claims are identified as those that have a diagnosis code that matches the codes in the typical Dx codes tabs (Typical Dx I-9 or Typical Dx I-10), or in the PAC Dx codes tab (PACs I-9 or PACs I-10) AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant admissions to acute and post-acute care facilities are also included in the denominator.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the ARR measure if they are considered not relevant to ARR care.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Denominator exclusions include exclusions of "patients" as well as "claims" not relevant to ARR care. Please refer to the enclosed excel workbook entitled (NQF_ARR_all_codes_risk_adjustment 06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
d. If the patient does not have at least 12 months of claims in the database (this helps eliminate incomplete episodes).

e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure if they meet one of the following criteria:
a. If none of the diagnosis codes on the claim are on the list of "triggers" or relevant diagnosis codes (either typical Dx or PAC Dx) for ARR.
b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for ARR.
c. If the ARR trigger hospitalization also triggers a major surgical procedure such as coronary bypass procedure or angioplasty, suggesting that ARR may be a comorbidity or an indication for the surgery.
d. The "principal" diagnosis on an inpatient stay claim during the episode time window triggers its own episode
e. The procedure code on a claim during the episode time window triggers its own episode

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
None

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Conceptual Model
Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, the remaining variance in PACs are due to factors that could be controlled by all providers that are managing or co-managing the patient, both during and after hospitalization.

Statistical Method:
Logistic Regression model to determine the probability of a patient incurring a PAC
Demographic variables, comorbid conditions, as well as clinical severity indicators are fed as independent risk factors into the model. Risk Factors are collected historically. Subtype information is collected from the index claim and any look-back period, if relevant. Subtypes are clinical severity indicators suggesting severity of the episode itself, for example, diagnosis of ventricular arrhythmias in a ARR patient. For each patient the "predicted" coefficients from the risk adjustment models are summed to give the predicted probabilities of the occurrence of a PAC.

Risk Factors :(Please refer to the enclosed excel workbook entitled (NQF_ARR_all_codes_risk_adjustment 06.30.15.xls). The risk factors along with their codes are listed in the tabs called "All Risk Factors I-9" and "All Risk Factors I-10" and also listed below:

AGE CONTINUOUS VARIABLE
GENDER FEMALE = 1 (MALE IS REFERENCE = 0)

Risk Factor #	Risk Factor Name
RF0101	Anoxic Brain Damage, persistent vegetative state
RF0102	Delirium, Meningitis, Encephalitis
RF0103	Previous Stroke, Paralysis
RF0104	Cerebral Palsy and Other Paralytic Syndromes
RF0105	Spinal Cord Disorders/Injuries
RF0106	Polyneuropathy
RF0107	Multiple Sclerosis
RF0108	Convulsions, Epilepsy
RF0109	Dementia
RF0110	Parkinson´s and Huntington´s Diseases

RF0111 Cerebrovascular Disease
RF0115 after care, rehabilitation
RF0201 visual loss, blindness, retinal tear, detachment
RF0301 ENT, Upper Respiratory Problems
RF0401 Respiratory Failure, O2, ventilator dependence
RF0402 Advanced COPD, Asthma
RF0403 Empyema, bronchiectasis, Pneumonias
RF0404 Aspiration Pneumonia, Laryngeal Problems
RF0406 TB, Pneumoconiosis, Aspergillosis
RF0407 Tobacco use, Lung disease due to External Fumes
RF0408 Other Lung Disease
RF0501 Previous Shock, Syncope, Vent Fibrillation
RF0503 Advanced CHF
RF0504 Cardiomyopathy, valve disorders
RF0505 Cardiac Arrhythmias, Heart Block
RF0506 Pacemaker, AICD
RF0507 Endocarditis, Other post surgical cardiac problems
RF0508 Other Cardiovascular Disease
RF0511 DVT, Pulm Embolism, Pulm Heart Disease
RF0512 Unstable Angina
RF0513 Hypotension, chronic, orthostatic
RF0514 Hyperlipidemia
RF0515 Intraaortic Balloon Pump
RF0516 ventricular assist device, ecmo, prolonged bypass
RF0517 Previous electrophysiology studies, cryoablation
RF0518 Recent AMI
RF0519 Previous PCI
RF0520 Previous CABG
RF0521 Previous Heart & Valve Surgery
RF0522 Previous aortic reconstruction
RF0523 Previos carotid endarterectomy
RF0524 Aortic and peripheral vascular disease
RF0525 Advanced Aortic and Vascular Disease
RF0601 GI Bleed
RF0602 Intestinal Obstruction/Perforation
RF0603 Acute Gastritis, Duodenitis
RF0604 Gastroduodenal Ulcer
RF0606 Intestinal Uro-genital Fistula
RF0607 Abdominal hernia w complications
RF0608 Vascular insufficiency of intestine
RF0609 Inflammatory Bowel Disease
RF0610 Irritable Bowel
RF0611 Diverticulitis, Meckel´s
RF0612 Digestive congenital anomalies
RF0613 Intestinal infection
RF0614 Esophageal Perforation, Hmg, Barretts, Compl Hiatal Hernia
RF0615 Abnormal weight loss
RF0616 Achalasia, Esophageal spasm, Stricture, Dysphagia
RF0617 GERD, Hiatal Hernia, Other Upper GI Disorders
RF0618 Previous Bariatric Surgery
RF0619 Hx of colon polyps, family Hx of colon cancer
RF0620 Enterostomy, GI devices, lap band
RF0701 Pancreatic Disease
RF0702 Perforation, fistula GB, bile duct, pancreas

RF0703 Gall stones, cholecystitis
RF0704 End-Stage Liver Disease
RF0705 Hepatitis, Cirrhosis, Other Hepatbiliary Disorders
RF0706 Recent Gall Bladder, Hepatobilary Surgery
RF0707 Acute Pancreatitis, pseudo cyst
RF0801 Bone/Joint/Muscle Infections/Necrosis
RF0802 Muscular Dystrophy
RF0803 Osteoporosis, ostetits deformans, pathological fracture
RF0804 Rheumatoid Arthritis and Inflammatory Connective Tissue Disease
RF0805 Gout and other crystal arthropathies
RF0806 Other arthropathies
RF0807 Osteoarthritis
RF0808 Joint Deformities
RF0809 Knee derangements
RF0810 Traumatic Dislocation Knee
RF0811 Dislocation Hip
RF0812 Synovitis, Ruture Tendon
RF0813 Status Knee Replacement
RF0814 Status Total Hip Replacement
RF0901 Decubitus Ulcer
RF0902 Skin and wound problems
RF1001 Diabetes, poor control
RF1002 Advanced diabetes
RF1003 diabetes
RF1101 Acute renal failure
RF1102 Dialysis Dependent
RF1103 Nephritis
RF1104 Chronic renal failure
RF1105 Urinary Tract Infections
RF1301 Endometriosis
RF1302 Fibroid uterus, benign tumors of female organs
RF1303 Pelvic Inflammatory disease
RF1304 Uterine prolapse, cystocele, vaginocele
RF1305 Female Harmonal Disorders
RF1306 Ovarian, Broad Ligament Disorders
RF1308 Other disorders of uterus, cervix
RF1309 Menopausal Disorders
RF1310 Menstrual Disorders
RF1401 Multiparity, multigravida
RF1402 Elderly Primi, other
RF1403 Poor obstetric history
RF1406 Cervical incompetence
RF1407 Abnormalities of uterus, female genital tract
RF1408 Hypertension, pre-eclampsia in Pregnancy
RF1409 Severe pre-eclampsia w HTN, Eclampsia
RF1410 Maternal, gestational diabetes, large for date
RF1411 Genital Herpes
RF1412 Infections of genitourinary tract, venereal disease in pregnancy
RF1413 Infectious Diseases in Mother
RF1414 Cardiovascular disease in Mother
RF1415 Mental Disorders in Mother
RF1416 Epilepsy in Mother
RF1417 Liver and biliary tract disorders in mother
RF1418 Kidney Disease in Mother

RF1419 Other Maternal conditions
RF1421 Cephalopelvic Disproportion due to maternal causes
RF1436 Peripartum Cardiomyopathy
RF1441 Previous Cesarean section
RF1450 Maternal Obesity, previous Bariatric Surgery
RF1454 Previous Rupture Uterus, Obstetrical Trauma
RF1458 Complicated Pregnancy Delivery
RF1460 Thrombophlebitis, DVT during Pregnancy
RF1461 Puerperal Sepsis, other major puerperal complications
RF1462 Obstetrical Embolism, Air, Amniotic Fluid, Pulm, Pyemic
RF1467 Tobacco Use in Mother
RF1601 Bleeding Disorders
RF1602 Severe Hematological Disorders
RF1603 Disorders of Immunity
RF1604 Nutritional and other Anemias
RF1605 Long-term use of anticoag, Aspirin
RF1701 Head and Neck Cancers
RF1702 Lung and Intrathoracic Cancers
RF1703 Neuroendocrine, Myeloproliferative Cancers
RF1704 Poorly differentiated, Secondary, Metastatic Cancers
RF1705 Other Tumors
RF1706 Acute Leukemia
RF1707 Cancer uterus, localized female organs
RF1708 Colorectal, Hepatobiliary and other GI cancers
RF1709 Breast, Prostate, Thyroid cancers
RF1710 Testicular Cancer and localized of male organs
RF1711 Cancer of Bladder and Urinary Tract
RF1712 Musculoskeletal Cancers
RF1801 Sepsis, MRSA, Opportunitistic infections
RF1901 Schizophrenia
RF1902 Major Depressive, Bipolar, and Paranoid Disorders
RF2001 Drug/Alcohol Psychosis
RF2002 Drug/Alcohol Dependence
RF2101 Drug Reactions, long term use of drugs
RF2102 Intra-abdominal injury
RF2201 Extensive Third-Degree Burns
RF2301 Major Organ Transplant Status
RF2302 Artificial Openings for Feeding or Elimination
RF2303 Complications of Medical & Surgical Care and Trauma
RF2304 severe morbid obesity
RF2305 morbid obesity
RF2306 obesity
RF2307 mild sleep apnea, hypoventilation
RF2308 moderate sleep apnea, hypoventilation
RF2309 obstructive sleep apnea
RF2310 Severe Protein-Calorie Malnutrition
RF2311 Mild-mod malnutrition
RF2401 Severe Head Injury
RF2402 Major Head Injury
RF2403 Vertebral Fractures without Spinal Cord Injury
RF2404 Falls, Fractures
RF2405 Amputation
RF2501 HIV/AIDS

Subtypes for ARR
Atrial Flutter / Fibrillation
Complication of Implanted device, graft
Electrophysiology Studies, Cryoablation
Heart Aneurysm and other Sequelae of AMI
Highgrade Heart Block
History of Sudden Death
Malfunction / Complication of Heart Device, H
Other Heart Blocks / Conduction Disorders
Sinus Node Dysfunction
Supraventricular Tachyarrhythmias
Ventricular Arrhythmias

The prevalence of the risk factors in our reference dataset are listed in the enclosed workbook entitled NQF_ARR_all_codes_risk_adjustment 06.30.15.xls – see tab "Risk Factor Prevalence". The output of the regression model are given in the same workbook in the tab 'Risk Model'.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Please refer to the enclosed excel workbook entitled (NQF_ARR_all_codes_risk_adjustment 06.30.15.xls).

Assembling the Denominator:

Using administrative claims database, patients with ARR are identified using one of two criteria: 1) Patients having an office visit with a trigger code of ARR in any position, followed by a second confirmatory office visit (with a trigger code of ARR in any position), at least 30 days apart, 2) Patients a Principal Dx of a ARR trigger code on an in-hospital stay claim. The trigger codes for ARR are provided in the tab called "Triggers I-9" or "Triggers I-10".

Patients are retained if they are 18 years of age or more, do not have a missing gender, have at least 12 month of claims in the database, have a maximum of 30-day enrollment gap for the entire episode time window, and have no outlier episode costs. All relevant professional, laboratory, imaging, ancillary and other claims that are incurred during the episode time window are included as part of the episode. Claims are considered relevant to ARR care if they have one of the diagnosis codes, as listed on the tab entitled Triggers I-9, Triggers 1-10, PACs I-9, PACs I-10, Typical Dx I-9, or Typical Dx I-10 in any position on the claim AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook.. Relevant admissions to acute and post-acute care facilities are also included in the denominator. All relevant pharmacy claims carrying codes that match the ingredients listed in the Pharmacy tab of the enclosed workbook are also included as part of the episode. Services are pulled as part of the ARR episode

based on the diagnosis codes as defined above or if they have a service code that is marked as "sufficient" for that episode.

If a patient has more than one concurrent episode, and the claim is relevant to both episodes, the claim could get multi-assigned, except in the case of procedural episodes that get carved out with respect to the index stay. So if an inpatient stay claim carried a principal Dx code that matched the trigger diagnosis code for ARR but they also had a procedure code for CABG (coronary artery bypass surgery), the stay claim would get uniquely assigned to CABG and not be counted with ARR.

Once all the episodes are assembled, episodes that match the exclusion criteria, such as those with outlier costs, are flagged (those with total episode costs less than 1st percentile or greater than 99th percentile), and excluded from the final analysis.

Assembling the Numerator:

For every episode included in the denominator, services are flagged as having a PAC (potentially avoidable complication) based on the criteria listed below:

Any service (professional, outpatient facility, ancillary) with a PAC code in any position on the claim

Any admission to an acute care facility, that is relevant to ARR

Relevant claims that do not have any PAC codes, and do not qualify as a PAC based on the criteria outlined above, are listed as typical claims. All included relevant pharmacy services are flagged as typical. Patients that have even a single PAC claim are counted as part of the numerator.

Calculating the measure:

Proportion of ARR patients that have PACs, is simply the ratio of patients with PACs within the HTN ARR population and is called the PAC rate as shown in the equation below:

PAC rate = Patients with ARR that have at least one PAC claim / Total number of ARR patients

A flow chart demonstrating the series of steps and the counts of patients at each step is shown in tab entitled Decision Tree of the enclosed workbook called NQF_ARR_all_codes_risk_adjustment 06.30.15.xls

Drill Down Calculations:

Further analysis from this construct helps create actionable reports.

For example as shown in the tab labeled PAC overview, not only do we have the PAC rate for a population, we can break them down by the PAC type – type 1 being directly related to ARR and so actionable by the servicing physician, while type 2 PACs are related to patient safety failures and can be improved by process improvement. Additionally, analyzing potentially avoidable admissions helps focus strategies in reducing them.

Risk Adjustment:

Once we have the observed PAC rates, we risk-adjust them for patient factors such as patient demographics, comorbidities collected historically, and for severity of illness or procedure using subtypes collected from the index stay and / or look-back period. This helps adjust for factors outside the providers control and levels the playing field for provider performance comparisons.

Unit of Analysis:
The unit of analysis is the individual episode.

Dependent Variable:
The dependent variable is a dichotomous variable indicating whether an episode had one or more claims assigned as a PAC (=1) or not (=0).

Independent Variables:
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_ARR_all_codes_risk_adjustment 06.30.15.xls

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., heart aneurysmobesity) or severity of the illness itself (e.g., unstable ventricular arrhythmiaangina). Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_ARR_all_codes_risk_adjustment 06.30.15.xls

As mentioned previously, to avoid creating perverse incentives all comorbidities and subtypes are identified prior to or at the very start of the episode. None are identified during the episode period.

Statistical Methods

We use logistic regression to model the probability of at least one PAC occurring during the episode. Only comorbidities and subtypes are included in the models as covariates if they are present in at least 10 episodes to prevent unstable coefficients. No further model building is conducted after the initial models are built. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but it does not make it a priority that all covariates in the model be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity of a episode condition, and lets each regression model determine for itself which of the factors are more significant for a specific episode. Non-significant covariates in episode models can not overly influence predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

When more than one line of business is included in the data, separate models are calculated for each sample (i.e., commercial, Medicaid etc.).

Provider Attribution and calculating PAC rates by provider:

Once episodes are constructed they are attributed to providers based on one of the various attribution rules. For ARR, episodes are attributed to the primary care physician or cardiologist with the highest count of office visits.

Using the logistic regression technique described above, a model is developed that gives estimates for each risk factor and subtype for the patients in the population analyzed. These estimates are used to develop patient-level probabilities for the occurrence of PACs. The patient-level probability estimates are summed to construct aggregated measures (e.g., facility/provider-level). This method is similar to the methods employed by the Centers for Medicare and Medicaid Services (CMS) and endorsed by the National Quality Forum (NQF) to construct similar facility- and practice-level measures (i.e., mortality, readmissions, etc.):

1. For each provider, the number of actual observed occurrences of the outcome is summed across all attributed patients with that episode, to give the observed PAC rates for the provider.
2. Similarly adjusted probabilities from the risk adjustment models are summed across all attributed patients to give expected PACs for the provider.
3. The observed sum is then divided by the summed probabilities (O/E). This number yields whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). This calculation yields a practice-level unstandardized performance ratio.
4. To facilitate accurate comparisons of rates across units of analysis, this ratio is then standardized to the community rate using the indirect method. Specifically, the provider-level rate is multiplied by the expected community rate, calculated as the sum of adjusted probabilities for every individual in the sample across all providers in the analysis. This measure, known as the standardized rate, represents what the unit's risk-adjusted rate would be for the outcome of interest if its patient population was reflective of the of the overall community.

The formula for this calculation is as follows:

Adj Outcome_j={(SUM Observed_ij)/(SUM Prob_ij)} × {(SUM Prob_i) / (# of episodes)}
Where individual is attributed to unit of analysis j (e.g., practice, provider, etc.)

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If patient related data is missing, the entire patient is excluded from the numerator as well as the denominator.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The information is based on a two-year claims database from a large regional commercial insurer. The database has over 3.2 million covered lives and $25.9 billion in “allowed amounts” for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

The methodology can be used on any claims database with at least two years of data and a minimum of 150 patients with the index condition or hospitalization. Having pharmacy data adds to the richness of the risk-adjustment models.
?The calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.
The methodology has been tested on databases of several health plans as well as on a few employer databases.

No data collection instrument was used.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual, Clinician : Team

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinician Office/Clinic, Other
If other: Across the care continuum

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2749_ARR_Testing_Reliability_Validity_HCI3.docx

NATIONAL QUALITY FORUM—Composite Measure Testing (subcriteria 2a2, 2b2-2b7, 2d)

Measure Number (*if previously endorsed*)**:** 2749
Measure Title: Proportion of Patients with Arrhythmias (ARR) that have a Potentially Avoidable Complication (during the episode time window)
Date of Submission: 06/30/15
Composite Construction:

☐ Two or more individual performance measure scores combined into one score
☐ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)
☐ Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be

demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. 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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The information is based on a two-year administrative claims database from a large regional commercial insurer. The database contains medical and pharmacy claims on over 3.2 million covered lives and more than $25.9 billion in "allowed amounts" for costs.

1.3. What are the dates of the data used in testing? April 1, 2012 – December 17, 2014

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.26***)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: Integrated Delivery System	☐ 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*)

There were a total of 6,728 providers in the data set. Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the reliability calculations. After this exclusion, there were 575 providers left.

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

After exclusions (see 2b.3.1 below), there were a total of 38,207 episodes of ARR included in the testing and analysis. Patients in these episodes were, on average, 49.1 years of age (range 18-64) and 54% were female. We did not have race information on these patients. All patients for this analysis met the trigger criteria of ARR as identified in our code tables and our programming rules.

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.

For the reliability analysis, we restricted the data to only providers with at least 10 attributed episodes. For risk adjustment, all episodes were used in the analysis, regardless of the provider to which they were attributed.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
None of the analyses included SDS variables.

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*)

We assessed the reliability of the measure to demonstrate that it sufficiently differentiates performance between providers using the beta-binomial method, which is applicable to measures of this type. Reliability is a measure that distinguishes the signal (the extent of performance variation between entities that is due to true differences in performance) from statistical noise. Our approach follows directly from the methods outlined in the technical report "The Reliability of Provider Profiling: A Tutorial" by J.L. Adams.

Reference:

Adams JL. The Reliability of Provider Profiling: A Tutorial. Rand Corporation. http://www.rand.org/pubs/technical_reports/TR653.html.

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*)

The table below provides a summary of the reliability score for different minimum sample size thresholds. For complete results, refer to the workbook entitled, NQF_ARR_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab to see provider-specific results.

Reliability Scores	**Minimum # Episodes Per Provider**		
	>=10	**>=25**	**>=50**
# of Providers (%)	575 (100)	232 (40)	103 (18)
Median (IQR)	0.66 (0.54,0.79)	0.80 (0.74,0.87)	0.88 (0.84,0.93)
Range	0.42-1.00	0.65-0.99	0.79-0.99

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

Reliability scores can vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or variation across patients within providers) whereas a reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (see Adams, 2009 cited above).

Although there was a wide range of scores across all providers with at least 10 episodes and scores for many were low, those among providers with at least 25 episodes were consistently good and continued to improve as provider sample size increased. This demonstrates that for providers with a minimum threshold number of episodes the measure sufficiently differentiates performance between them.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b2. VALIDITY TESTING

***Note**: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

2b2.1. What level of validity testing was conducted?

☐ **Composite performance measure score**

- ☐ **Empirical validity testing**
- ☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)
- ☐ **Systematic assessment of content validity**

☐ **Validity testing for component measures** *(check all that apply)*

***Note**: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

- ☐ **Endorsed (or submitted) as individual performance measures**
- ☐ **Critical data elements** (*data element validity must address ALL critical data elements*)
- ☐ **Empirical validity testing of the component measure score(s)**
- ☐ **Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. 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)*

Content validity was built into the development of the definitions of potentially avoidable complications (PACs). This involved working with clinicians who are experts in their respective fields and specific to the episodes for which PACs are being measured. In particular, the clinical experts focused on whether or not a potentially avoidable complication can be deemed as such for a specific episode of care, and help

defined and review all of the diagnosis and procedure codes for each PAC. The enclosed link lists clinicians who have participated in the various Clinical Working Groups (http://www.hci3.org/content/clinical-working-group-contributors). Some of the clinical experts have also participated in monthly webinars that highlight the clinical aspects of these measures (http://www.hci3.org/content/using-ecrs-providers).

In addition, we illustrate that our measure has face validity in several ways.

Beyond the up front work performed by clinical experts, the validity of the measure has also been tested in various real world settings. For example, we have presented results of claims data analyses that reveal the frequency and costs of PACs to physicians in several different healthcare systems involved in our pilot site implementations, as well as to medical directors from the employer coalitions and the health plans that provided the dataset to run the analyses. Some of these implementations include the Pennsylvania Employee Benefits Trust Fund and local provider groups and hospital, Horizon Blue Cross Blue Shield of NJ and many physicians and health systems.

In addition, we have performed dozens of analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:

- Vermont Payment Reform Commission
- Maine Health Management Coalition
- WellPoint / Anthem CT
- NY State Medicaid
- CT Medicaid
- CO All-payer Claims Database, Center for Improving Value in Health Care

These analyses and their results have influenced, and continue to influence, the development of various public reporting, payment reform and delivery system reform efforts. To-date, we have never experienced either wholesale or partial rejection of the results of analyses showing rates of PACs, which demonstrates the level of acceptability – face validity – of the measures from the payer, policymaker, employer and payer communities.

As importantly, measures of potentially avoidable complications have face-validity with consumers. In a series of focus groups, Judy Hibbard and colleagues[1] examined the impact of presenting information about price and quality of certain providers in influencing the decisions of consumers. They tested the validity of PACs as a discriminator of quality, as well as other measures of quality, and used the dollar symbol to illustrate the level of price, much like is done for restaurant reviews. When the PAC measure was used, respondents selected the providers with the lowest PAC rates with a high level of confidence in choice, and used it as a surrogate for a strong quality signal. To the contrary, when more standard measures of quality were used, consumers tended to ignore them and use price as a surrogate for quality. As such, what the researchers found is that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer “defects” – and lower price.

Finally, our measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs)[2], Hospital Inpatient Quality Reporting measures [3], Avoidable Readmissions [4,5], AHRQ defined patient safety indicators (PSIs) [6], NQF endorsed patient safety

measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates [7].

References:

[1] Hibbard JH, Greene J, Sofaer S, Fiminger K, and Hirsh J. An Experiment shows that a well-designed report on Costs and Quality can help consumers choose High-Value Health Care. *Health Affairs* 2012; 31(3): 560-568. doi: 10.1377/hlthaff.2011.1168

[2] CMS defined Hospital Acquired Conditions: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

[3] CMS operated Hospital Inpatient Quality Reporting Program: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

[4] Jencks SF, Williams MV, and Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. N Engl J Med 2009 (Apr); 360 (14): 1418-1428. doi: 10.1056/NEJMsa0803563.

[5] Casalino LP, Pesko MF, Ryan AM et.al. Small Primary Care Physician Practices have low rates of Preventable Hospital Admissions. Health Affairs, 2014; 33(9): 1-9. doi: 10.1377/hlthaff.2014.0434.

[6] Agency of Healthcare and Quality defined Patient Safety indicators: http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx

[7] NQF endorsed measures: Quality Positioning System: http://bit.ly/1E5ZdP7

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Not applicable.

2b2.4. 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?*)

Given the significant clinical input that went into developing the measure, the widespread use and acceptance the measure has gained among a wide variety of individuals and organizations across the health system (public and private payers, clinicians, consultants, patients, etc.) [1-13], and the parallels between this measure and other measures that are in widespread use, this demonstrates that the measure has strong face validity.

References:

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)

2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.

3. de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.

4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.

5. Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.

6. Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.

7. François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)

8. de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

9. Rastogi A, Mohr BA, Williams JO, Soobader MJ, de Brantes F. Prometheus Payment Model: Application to Hip and Knee Replacement Surgery. Clin Orthop Relat Res 2009; 467(10): 2587-2597.

10. de Brantes F and Rastogi A. Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. The Commonwealth Fund 40, publ. 2008; 1146:1-14.

11. de Brantes F, Gosfield A, Emery D, Rastogi A and G. D'Andrea, "Sustaining the Medical Home: How Prometheus Payment Can Revitalize Primary Care", Robert Wood Johnson Foundation Report, May 2009, http://www.rwjf.org/pr/product.jsp?id=42555, accessed October 2009.

12. de Brantes F, Camillus J. Evidence-informed case rates: a new health care payment model [Internet]. New York (NY): Commonwealth Fund; 2007 Apr [cited 2007 May 20]. Available from: http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=478278, Accessed Aug 1 2013.

13. Satin DJ, and Miles J. Performance Based Bundled Payments: Potential Benefits and Burdens. Available from: http://student.med.umn.edu/p4p-new/sites/default/files/MN%20Med%20Bundles%20Special%20Report%20-%20Satin.pdf, Accessed Aug 1 2013.

2b3. EXCLUSIONS ANALYSIS

Note: *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*

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

2b3.1. 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)

No formal exclusion testing was done since no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers.

Exclusions include exclusions of "patients" as well as "claims" not relevant to ARR care.

1. "Patients" are excluded from the measure if they meet one of the following criteria:
a. If age is < 18 years
b. If gender is missing
c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
d. If the patient does not have at least 12 months of claims in the database (this helps eliminate incomplete episodes).
e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure if they meet one of the following criteria:
a. If none of the diagnosis codes on the claim are on the list of "triggers" or relevant diagnosis codes (either typical Dx or PAC Dx) for ARR.
b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for ARR.
c. If the ARR trigger hospitalization also triggers a major surgical procedure such as coronary bypass procedure or angioplasty, suggesting that ARR may be a comorbidity or an indication for the surgery.
d. The "principal" diagnosis on an inpatient stay claim during the episode time window triggers its own episode
e. The procedure code on a claim during the episode time window triggers its own episode.

2b3.2. 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*)

We started with a total ARR population of 81,216 episodes. After all the exclusions were applied, the remaining ARR population included in the analysis consisted of 38,207 episodes. As mentioned above, no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers. As such, no formal exclusion testing was done.

2b3.3. 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*)

No formal analysis was done on the impact of exclusions on performance scores.

Descriptive Explanation:

Exclusions of patients were for the following reasons. Some are for comparative purposes and some for medical

reasons.

(a) Comparative Purposes:
We excluded patients that did not have complete enrollment for the entire episode time window. This was done to ensure that the database had complete information on patients to be able to create the entire episode. Including patients with only a partial episode window could distort the measure by artificially reducing the actual count of patients with PACs.

(b) Medical Reasons:
Patients with outlier costs (less than 1st percentile value or greater than 99th percentile) were considered to be different from the general pool, and excluded from both the numerator and the denominator. This is another way to ensure that episodes are complete (because incomplete episodes may have very low costs), and do not bring in random noise into the analysis due to inappropriate codes or services (high outliers).

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
Note: *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used? *(check all that apply)*
☐ **Endorsed (or submitted) as individual performance measures**
☐ **No risk adjustment or stratification**
☐ **Statistical risk model** with 170 potential risk factors and episode specific subtypes
☐ **Stratification by risk categories**
☐ **Other,** Click here to enter description

2b4.2. 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

Risk factors are comorbidity indicators collected from historical claims before the start of an episode. These are universally applied across all episodes. This list was selected based on input from clinical experts in clinical working groups. In addition, the Clinical Working Groups identified episode specific severity markers that were called episode subtypes and they help distinguish an episode as being more severe than another.

All risk factors and subtypes must be present prior to, or at the start of the episode and are identified using diagnosis codes in the patient's historical claims.

To be included in the risk adjustment models, any risk factor or subtype must be present in at least 10 episodes. Beyond this no further model building was conducted to add or remove risk factors or subtypes from the model after it was initially run. This reflects a desire to explain as much variation in the probability of having a PAC

as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates cannot overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

2b4.4. What were the statistical results of the analyses used to select risk factors?

As explained above, no formal analysis was conducted to select risk factors. In fact, all potential risk factors and subtypes with a count of at least 10 episodes were retained to serve as predictors. The goal was to achieve a more complete explanatory model rather than achieve parsimony.

Please reference the tabs titled Risk Model in the NQF_ARR_all_codes_risk_adjustment_06.30.15.xls workbook to see the list of risk factors that met the selection criteria.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Not Applicable since our analysis did include SDS variables

2b4.5. 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*)

Model Development Approach

We used logistic regression to model the probability of at least one PAC occurring during the episode. The model included all covariates that were identified through the process above. No further model building was conducted after the initial model was run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates can not overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

For a more complete description of the risk adjustment approach, please see the document entitled, "PACs and Severity Adjustment Fact Sheet" that accompanies this submission.

Approach to Model Testing and Validation

To determine the validity and performance of the model, we used the split sample method to divide the patient sample randomly into: 1) the model building data set (80% of the sample) and 2) the test data set (20% of sample. The model was built using logistic regression on the first data set and then the coefficients from the development model were tested in the second dataset. Area under the curve (AUC) and the c-statistic were used to compare the predictive ability of the model in each of the data sets. Hosmer-Lemeshow Goodness-of-Fit tests and comparisons of observed to expected probabilities across risk deciles were further examined to assess the model's overall predictive accuracy.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

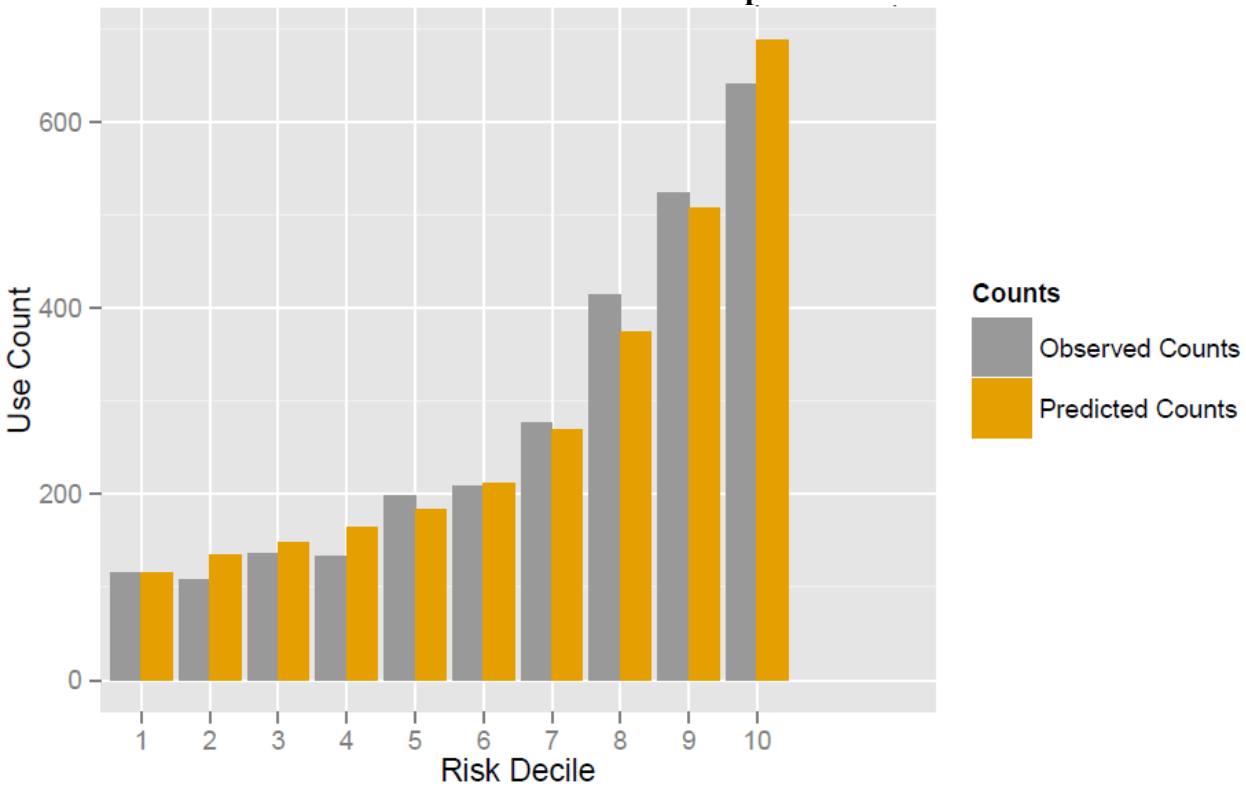
Sample	Accuracy (%)*	AUC
Test	75.2%	0.781
Validation	75.0%	0.773

*Episodes with predicted probabilities <50% were classified as having a predicted 0 (not having a PAC). Episodes with predicted probabilities >50% were classified as having a predicted 1 (having a PAC)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Sample	Chi Square	Degrees of Freedom	p-value
Test	202.2	8	<0.0001
Validation	58.9	8	<0.0001

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:



2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The C statistic is a measure of the extent to which a statistical model is able to discriminate between a patient with and without an outcome. The c-statistic ranges from 0.5 to 1.0. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong [1]. The purpose of the model is to adjust for patient-related factors. The remaining unexplained differences in PAC rates are due to factors that could be controlled by all providers that are managing or co-managing the patient, during the entire episode time window.

The c-statistics of the testing and validation samples (0.781 and 0.773, respectively) indicate that the risk models have strong discriminatory power. Indeed, the accuracy values show that the model correctly predicts whether an episode had or did not have a PAC about 75% of the time, well above what would be expected if the predictions were made at random (i.e., 50%). Although the H-L test was significant for the testing sample, meaning that the model is not a good fit, this test is generally known to be sensitive to the number of groupings used and sample sizes. Nevertheless, the risk decile plot plot shows that the model predicts PACs similarly to the number of observed PACs across all deciles.

Overall, the results of the risk adjustment analysis strongly indicate that the model has sufficient predictive power.

Reference:
[1] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Edition)*. New York, NY: John Wiley & Sons; 2000.

2b4.11. 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*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To directly compare PAC rates across providers while also appropriately accounting for differences in patient severity, we calculated a risk-standardized PAC rate for each provider. This method is similar to calculations used by others for reporting outcomes measures [1]. For each provider, the ratio of observed attributed episodes with PACs to the expected number of attributed episodes with PACs given the patient's risk factor and estimated from the risk-adjustment model was calculated. This number yielded whether the provider had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). We then multiplied this ratio by the overall expected PAC rate across all providers to obtain the risk-standardized PAC rate for the provider. This measure represents what a provider's PAC rate would be if its patient population was reflective of the overall population.

Because providers with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. Comparison of risk-adjusted PAC rates gives a measure of the provider's relative performance. Our analysis compared risk-standardized PAC rates across providers. We analyzed various descriptive statistics including the range in PAC rates, medians, interquartile range, etc.

References:

[1] See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

2b5.2. 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*)

Summary of Unadjusted and Adjusted Performance Scores Across Providers

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25
Unadjusted		
Median (IQR)	36% (26%, 45%)	34% (28%, 44%)
Range	0%-100%	6%-98%
Adjusted (RSPR)*		
Median (IQR)	36% (29%, 42%)	36% (31%, 41%)
Range	0%-92%	8%-67%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_ARR_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.e., *what do the results mean in terms of statistical and meaningful differences?*)
The variation in risk-adjusted rates suggests there are meaningful differences in performance between providers in risk-standardized PAC rates for patients with an episode of ARR.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This criterion is directed 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 SDS 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.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

2b6.2. 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*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

If patient related data is missing, the entire patient is excluded from the numerator as well as the denominator.

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Additionally, if a patient had one or more PACs, it is simply counted as a 1, i.e., flagged as having a PAC.

According to our measure definition, in constructing the measure it is possible for a provider to have only one or some types of PACs and not others. Alternatively, the provider may have all PAC types occur for their

patients. The measure only considers whether any PAC occurred regardless of the type, and all PAC types are weighted equally, therefore we believe, there is no potential for the absence of specific PAC types to bias performance scores for individual providers.

For these reasons, no formal analyses were done on missing data.

2b7.2. 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*)
Not applicable

2b7.3. 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*)

Not applicable

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH
Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook entitled NQF_ARR_all_codes_risk_adjustment_06.30.15.xls provides outputs from empirical analysis. The tab labeled "PAC overview" demonstrates percentage of episodes that had at least one PAC, and provides the breakdown of PACs: 1) by the type of PAC whether directly related to index condition or due to patient safety failures; 2) the setting of the PAC, whether seen in the in-patient setting, out-patient facility or during professional visits; and 3) preventable hospitalizations.

The "PAC Drill Down Graph" provides further detail on each component of the PAC and their frequency. As can be seen by the individual counts and the graph, while each individual PAC may have such small occurrences that no meaningful comparisons in provider performances could be made; together, they add value to provide a comprehensive picture that result in meaningful numbers. The aggregation of PACs to a comprehensive, composite measure, in itself provides the parsimony that is so desirable.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

All PACs, as clinically defined by the subject matter experts were used with equal weighting. Since the emphasis of the PAC measure is to identify the occurrence of PACs in any setting, a simple and straightforward approach was adopted.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc*.; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

No formal analysis was performed.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.e., *what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Additionally, if a patient had one or more PACs, it is simply counted as a 1, i.e., flagged as having a PAC. The measure only considers whether any PAC occurred regardless of the type, or the site, and all PAC types are weighted equally. Therefore, no formal analysis of individual components was performed.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

We chose not to weight the components of the measure.

Considerations were given to the fact that preventable hospitalizations may be given more weight, than PACs identified in a doctor's office. Similarly PACs in an in-patient setting may have more serious implications on a patient's ultimate outcome, than PACs occurring in an outpatient setting. Additionally, preventable hospitalizations as well as index hospitalizations, each with longer lengths of stay, may have serious PACs. But how do we weigh these effects? An alternative model was considered, where cost could be considered as a surrogate for the weights. Higher cost PACs could imply more serious PACs. However, differences in costs could be driven by many issues other than the PAC itself, such as unit price of the service, method of reimbursements, contracting arrangements etc.

Furthermore, in-patient facility billing does not allow for the distinction of PAC related costs from other costs within the stay. We would fail to capture PAC related costs within the stay and potentially underweight those. As a result, the decision was made to avoid weighting and keep the measure as a straightforward count.

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

Measuring all providers with the same yardstick will provide consistent results and reasonable comparisons over time. If the goal is to reduce PACs, then the PAC measure as was constructed with the help of various experts in the field would provide reasonable comparisons. A word of caution however pertains to the sample size of the provider panel before making any reasonable conclusions.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved for high reliability in one dataset would apply to another.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

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

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
As part of our general implementation of these measures and related analyses, we have worked through dozens of different and sometimes very large datasets. From Medicare to Medicaid to regional and national commercial carriers, as well as individual employers, the principal lesson learned is the heterogeneity of the data sets and the significant variability in fill rate of critical data elements. As a result, we have created highly specific recommendations for which data elements are required to ensure measure validity, the accuracy of those data elements, and their completeness in the dataset. When claims datasets are organized in the way we specify in the measure analysis, and contain the coding information required, the analysis of the measure and its results are highly reliable.

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 calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Payment Program Blue Cross Blue Shield of North Carolina Blue Cross Blue Shield of New Jersey Pennsylvania Employee Benefits Trust Fund https://www.bcbsnc.com/ http://www.horizonblue.com/ https://www.pebtf.org/ Quality Improvement (Internal to the specific organization) Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

4a.1. **For each CURRENT use, checked above, provide:**

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

Measures associated to potentially avoidable complications (PACs) are in use today with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system (public and private payers, clinicians, consultants, all-payer claims database stewards, etc.) [1-8]. They are being used in various capacities in different pilot site implementations. To name a few:

•BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
•BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers

In addition, the PAC measures are incorporated by the following organizations in their bundled payment programs:

•BCBSSC – for CABG and PCI programs
•Horizon BCBSNJ– for CHF and CABG programs
•BCBSNC
•PEBTF in PA

http://www.ajmc.com/interviews/Lili-Brillstein-on-How-Bundled-Payments-Are-Tranforming-Healthcare

In these programs they look at PACs related to the measure for process improvement activities and for practice re-engineering.

We have created reports for rates of PACs for the following organizations:
-Vermont Payment Reform
-Maine Health Management Coalition
-WellPoint / Anthem CT
-NY State Medicaid
-CT Medicaid
-CO All-payer Claims Database, Center for Improving Value in Health Care

There are several companies that are leveraging these measures to create analytics and software for customers – these include HealthQx, Aver Informatics, McKesson, and TriZetto.

Below are some references that highlight our work with Potentially Avoidable Complications (PACs):
1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)
2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.
3.de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.
4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.
5.Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.
6.Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.
7.François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)
8.de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

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

N/A

4a.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.*)

Measures associated with PACs are currently in use as described in the prior section. In addition, we are working with several not-for-profit and for-profit organizations to provide them with the algorithms needed to calculate rates of potentially avoidable complications. Some of these organizations include:

Fair Health – based in NY and whose mission is to increase transparency of provider cost and quality,

CastLight – based in CA and serving large employers. We currently provide CastLight with Bridges To Excellence recognitions and will work with them to augment provider transparency by using PAC measures,

MA APCD (Massachusetts All Payers Claims Database) Council – we currently have an agreement in place with the MA APCD Council to produce PAC measures on hospitals and physicians and report back to the council with tests of reliability and validity of the measures. The purpose is to authorize the publication of these measures,

Maryland Health Care Cost Commission – we have a two year agreement to produce measures of cost and quality for public dissemination.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

We do not have any public information to share about the improvements in rates of potentially avoidable complications, as the implementation of these measures is too recent to provide valid comparisons. Further, some of the definitions of PACs have changed since the measures were initially endorsed, making comparisons even more difficult and unreliable.

Nevertheless, the variation in performance scores presented in Section 1b.2 indicates that there are differences between providers in their risk-adjusted PAC rates (higher scores equal worse performance). This suggests that real opportunities exist to identify lower performing providers and reduce the overall occurrence of PACs.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

Performance results provide summary PACs rates by provider, which can be used by payers and providers in a number of ways to improve the quality of care.

From the payer perspective, payers can use this information to 1) create a high-value provider networks, 2) work with high-value providers to share best practices, 3) incentivize low-value providers to improve, 4) modify their insurance design to activate consumers to select the right care from the right providers at the right time.

From the provider perspective, providers can 1) view services and activity for their patients longitudinally across the entire care continuum, such as frequency of readmissions and ED visits and drill down on patients with high PAC rates, 2) review actionable drill down reports to identify the most frequent PACs across all patients to create care pathways and process improvement plans to impact the most frequent PACs.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

No unintended consequences were reported, but there is the potential for:
1. Under-coding of PACs in the claim stream resulting in under-reporting the actual rate and/or providers gaming the measures
2. Payers calculating the measures even with inadequate sample sizes and using the results to penalize providers

The measure is designed for transparency efforts and to spur quality improvement. Detailed PAC reports can help providers identify areas of quality improvement. Even detailed reports of small samples of patients can be helpful for quality improvement purposes, but not for public reporting. To mitigate the potential for invalid provider comparisons, we specify in this submission the minimum sample size needed to ensure the reliability of a provider's score. Ultimately, there isn't any good way to prevent provider gaming of the measure by under-coding claims, however, under the current DRG payment methodology, many providers would be penalized by under-coding PACs since these codes often result in the assignment of more complicated DRGs.

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)**

0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

5a. Harmonization

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 completely harmonized?

No

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

Some of the measures listed in the prior section are, fully harmonized with the submitted measure, in particular, 0705, 0708, and 0709. Other measures such as 0337 and 0450 are in fact, subsets of our measure. However, there are some measures that are not harmonized, in particular the 30-day all-cause readmission measures and the Hospital wide all-cause readmission measure. While the submitted PAC measure include hospitalizations and readmissions that occur during the episode time window, the hospitalizations, by definition, have to be relevant to the underlying condition. For chronic conditions, most relevant hospitalizations within the entire episode time window are considered potentially avoidable. PACs include readmissions and are designed to enable accountability at the locus of provider control as well as some shared accountability between settings, centered around a patient, and for a specific medical episode of care. In that sense, they are consistent with the all-cause 30-day readmission rates, but represent a subset of those admissions. However, they do extend to the entire episode time window. As such, the PAC measures, as submitted, don't create added burden of reporting because the readmissions reported are simply a part of the broader 30-day all-cause readmission measures already endorsed by NQF. Because PAC measures are comprehensive, they include patient safety events as well as other adverse events, including hospitalizations and ED visits during the entire continuum of care. As a result, they are a comprehensive measure of avoidable complications for a specific medical episode. The data collection for all of the HCI3 measures is automated by a software package and is fully harmonized with all other PAC measures. A single download automates creation of all reports related to each of the PAC measures.

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.)

The PAC measure is a comprehensive measure representing "all-cause harms". It looks at all potentially avoidable complications in patients hospitalized with AMI during the stay or for 30-days post-discharge. It looks at readmissions, emergency room visits, adverse events due to errors of omission or commission. It looks at complications that are due to patient safety failures, and also those directly related to the index condition. These are a cause of significant waste and quality concerns for patients with an AMI episode. As such, the measure can provide clinicians with an overall and comprehensive view, in one measure, of all potentially avoidable complications for a patient and drive quality improvement efforts.

For clinicians and facilities increasingly engaged in value-based payment efforts and/or driving quality improvement for population health, the value of a PAC measure over a series of related, but more discrete measures, is that one can better determine if the sources of complications primarily stem from activities within the facility or outside the facility, and the specific nature of the complications that have a higher frequency of occurrence. For providers, it's far easier to construct a quality dashboard from a parsimonious set of measures, and that's what PAC measures offer.

Further, as a comprehensive outcome measures, PACs are also useful for public transparency of quality, as substantiated by the research from Judy Hibbard and colleagues previously cited in the "testing" section of this submission.The PAC measure is a comprehensive measure representing "all-cause harms". It looks at all potentially avoidable complications in patients hospitalized with AMI during the stay or for 30-days post-discharge. It looks at readmissions, emergency room visits, adverse events due to errors of omission or commission. It looks at complications that are due to patient safety failures, and also those directly related to the index condition. These are a cause of significant waste and quality concerns for patients with an AMI episode. As such, the measure can provide clinicians with an overall and comprehensive view, in one measure, of all potentially avoidable complications for a patient and drive quality improvement efforts.

For clinicians and facilities increasingly engaged in value-based payment efforts and/or driving quality improvement for population health, the value of a PAC measure over a series of related, but more discrete measures, is that one can better determine if the sources of complications primarily stem from activities within the facility or outside the facility, and the specific nature of the complications that have a higher frequency of occurrence. For providers, it's far easier to construct a quality dashboard from a parsimonious set of measures, and that's what PAC measures offer.

Further, as a comprehensive outcome measures, PACs are also useful for public transparency of quality, as substantiated by the research from Judy Hibbard and colleagues previously cited in the "testing" section of this submission.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PACs_and_Severity_Adjustment_Fact_Sheet_HCI3-635719689859912707.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.2 **Point of Contact:** Francois, dr Brantes, francois.debrantes@hci3.org, 203-270-2906-
Co.3 **Measure Developer if different from Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.4 **Point of Contact:** Amita, Rastogi, Amita.rastogi@hci3.org, 213-934-9624-

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.
From 2006 onwards, and under the auspices of various funding organizations, HCI3 has convened and managed, or helped to convene and manage, Clinical Working Groups to inform the development and refinement of the measures. For example, in 2011, 2012 and 2013, HCI3 worked collaboratively with the American Board of Medical Specialties and the American Medical Association's Physicians Consortium for Performance Improvement, under a federal contract, to convene and get input from various clinical experts on definitions of episodes of care and their sequelae, including avoidable complications.

Some of the clinical experts that have contributed to our work include:
-Dr. John Allen, American Gastroenterology Association (AGA)
-Dr. Morton Arnsdorf, Cardiologist, University of Chicago, IL
-Dr. Peter Bach, Memorial Sloan Kettering Cancer Center (MSKCC)
-Dr. Peter Basch, Primary Care, Medstar Health, DC
-Dr. Justin Beckelman, Radiation Oncology, University of Pennsylvania, PA
-Dr. Debra Bingham, Executive Director, California Maternal Quality Care Collaborative (CMQCC) at Stanford University, CA
-Dr. John Birkmeyer, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Linda Bosserman, Wilshire Oncology Medical Group, CA
-Dr. Matthew Brengman, American Society of Metabolic and Bariatric Surgery (ASBMS)
-Dr. Joel Brill, American Gastroenterology Association (AGA)
-Dr. George Cautilli, Cautilli Orthopedic Surgical Specialists PC, Yardley, PA
-Dr. Ashwini Davison, Internist, Johns Hopkins Hospital, MD
-Dr. James Denneny, III, American Academy of Otolaryngology – Head and Neck Surgery (AAO-HNS)
-Dr. Chris Gallagher, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Robert Haralson, III, American Academy of Orthopedic Surgeons (AAOS)
-Ms. Dawn Holcombe, Executive Director, Connecticut Oncology Association, CT
-Dr. Colin Howden, American Gastroenterology Association (AGA)
-Dr. John Knightly, American Association of Neurological Surgeons (AANS)
-Dr. Larry Kosinski, American Gastroenterology Association (AGA)
-Dr. Nalini Krishnan, Obstetrics & Gynecology, MN
-Dr. Kelly Kyanko, Internist, NYU School of Medicine, NY
-Dr. Tara Lagu, Internist & Infectious Disease, Baystate Medical Center, MA
-Dr. Robert Lee, Society of Thoracic Surgeons (STS)
-Dr. Alex Little, Society of Thoracic Surgeons (STS)

-Dr. Michael London, Orthopedic Surgeon, OMNI Orthopedics, OH
-Dr. Elliott Main, Obstetrics & Gynecology, California Pacific Medical Center, CA
-Dr. Constantine Mantz, 21st Century Oncology, FL
-Dr. Joseph Messer, Cardiologist, Rush University Medical Center, IL
-Dr. David Metz, American Gastroenterology Association (AGA)
-Dr. Ronald Nahass, Infectious Disease Care, NJ
-Dr. Ajay Nehra, Urologist, Rush University Medical Center, IL
-Dr. Francis Nichols, Society of Thoracic Surgeons (STS)
-Dr. Patrick O'Connor, Primary Care, HealthPartners, MN
-Dr. Sara Perkel, National Comprehensive Cancer Network, PA
-Dr. David Peura, American Gastroenterology Association (AGA)
-Dr. John Ratliff, American Association of Neurological Surgeons (AANS)
-Dr. Steven Schutzer, Connecticut Joint Replacement Institute, CT
-Dr. Leif Solberg, Primary Care, HealthPartners, MN
-Dr. Scott Sporer, Midwest Orthopedics at Rush, Chicago IL
-Dr. Bonnie Weiner, Cardiologist, Worcester Medical Center, MA
-Dr. Jonathan Weiner, Bariatric Surgery codes, Prof of Health Policy and Management, Johns Hopkins University, MD
-Dr. Janet Wright, Cardiologist, Northstate Cardiology Consultants, CA

Measure Developer/Steward Updates and Ongoing Maintenance

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

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

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

Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

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 #: 2751

De.2. **Measure Title:** Proportion of Patients undergoing an Angioplasty Procedure (Percutaneous Coronary Intervention - PCI) that have a Potentially Avoidable Complication (during the episode time window)

Co.1.1. **Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)

De.3. **Brief Description of Measure:** Percent of adult population aged 18 + years who had a percutaneous coronary intervention (PCI) procedure, are followed for at least 90-days, and have one or more potentially avoidable complications (PACs). PACs may occur during the index stay or during the 90-day post discharge period.

Please reference attached document labeled NQF_PCI_all_codes_risk_adjustment_06.30.15.xls, in the tabs labeled PACs I-9 and PAC I-10 for a list of code definitions of PACs relevant to PCI.

We define PACs as one of two types:

(1) Type 1 PACs - PACs directly related to the index condition: Patients are considered to have a PAC, if they receive services during the episode time window for any of the complications directly related to PCI, such as for hypotension, cardiac arrest, fluid and electrolyte disturbances etc.

(2) Type 2 PACs - PACs suggesting Patient Safety Failures: Patients are also considered to have a PAC, if they receive services during the episode time window for any of the complications related to patient safety failures such as for sepsis, infections, phlebitis, deep vein thrombosis, pressure sores etc.

All readmissions in a patient with PCI are considered potentially avoidable and flagged as PACs.

PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook labeled NQF_PCI_all_codes_risk_adjustment_06.30.15.xls serves as an example. The tab labeled PAC overview gives the percent of PCI episodes that have a PAC and the tab labeled "PAC drill down" gives the types of PACs and their frequencies in PCI episodes within this dataset.

The information is based on a two-year claims database from a large regional commercial insurer. The abase had over 3.2 million covered lives and over $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

1b.1. **Developer Rationale:** Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 40% of its plan members with hypertension

incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow- up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.

4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

6) BCBSNC: Blue Cross Blue Shield of North Carolina:
https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5.
http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/

11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.

14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

S.4. **Numerator Statement:** Number of patients who underwent a percutaneous coronary intervention (PCI) procedure, are followed for at least 90-days, and have one or more potentially avoidable complications (PACs) during the episode time window.
S.7. **Denominator Statement:** Adult patients aged 18 years and above who underwent an Angioplasty (percutaneous coronary intervention - PCI) procedure and are followed for at least 90-days
S.10. **Denominator Exclusions:** Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the PCI measure if they are considered not relevant to PCI care.

De.1. **Measure Type:** Composite, Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Clinician : Group/Practice, Clinician : Team, Facility, Health Plan, Population : National, Population : Regional, Population : State
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

IF this measure is included in a composite, NQF Composite#/title: n/a – 2751 is an "any or none" composite measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)
Component Measures (if endorsed or submitted for endorsement: n/a. The individual complications are considered measurable components.

Is this a MAINTENANCE measure submission? ☐ **Yes** ☒ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: n/a **Most Recent Endorsement Date:** n/a

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating clinical evidence asks if the there is a relationship between the measured health outcome and at least one clinical action is identified and if it is supported by the stated rationale. For a composite measure, the developer must discuss the reasoning for the composite quality constructs, the rationale for constructing, & aggregation and weighting of measure components.

- This new risk-adjusted (by age, gender and clinical co-morbidities) outcomes measure that assesses the proportion of adult patients with claims triggered by undergoing Angioplasty Procedure (Percutaneous Coronary Intervention - PCI) with at least one Potentially Avoidable Complications (PAC) within 12 months of PCI triggered claims data.
- Based on NQF's criteria, this measure is considered an "any or none" composite measure that assesses if 1 or more PACs or "care defects" have occurred for the index episode. For this composite measure, the individual complications considered the measurable components. PACs are classified in two types: 1) related to PCI, and 2) related to Patient Safety Failures, combining the 2 types into a single PAC rate to calculate the proportion of patient with 1 or more PAC. PACs are considered unwarranted health outcomes that are measured during an index condition for use at the facility, provider system or purchaser/payer levels to identify quality of care gaps between practices and hospitals.
- The developer links primary & secondary prevention care gaps, poor patient education, poor care coordination and poor follow-up increase unnecessary ER visits, hospitalizations, readmissions, and mortalities to increased PACs, and state that PACs for PCI patients should occur rarely in well-managed patients, providing potential avoidable PCI complications, that include (but not limited to) peri-procedural bleeding, emergent CABG, and death.
- The evidence for Patient Safety Failure PACs is described to be within the influence of the measured entity, though the rationale for selecting some of the identified PACs is not clear (e.g., post procedural fever, oral bisphosphonates, hallucinations). The developer provides an extensive list of comorbidities as risk factor for increased PAC potential, though the severity is not captured in consistently within the claims data.
- In addition to linking processes of care to outcomes, the developer provides an extensive PAC literature review in sections 1a.2. and 1a.2.1. for PCI, Patient Safety Failures & processes of care, as well as background information on the process for PAC development.
- The developer discusses the rationale for constructing, aggregation and equal weighting for the measure.

Questions for the Committee:

- *Does sufficient evidence exist connecting Patient Safety Failures to the PCI index episode?*
- *Does sufficient evidence exist between the measured health outcome and at least one clinical action identified and supported by the stated rationale?*

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

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

- PCI PCA performance gap data are calculated from PROMETHEUS administrative claims data from April 1, 2012 through December 17, 2014, for 5,898 of 10,177 (58.0%) PCI episodes (in 3,258,706 unique beneficiaries) and 41 of 565 (7.1%) facilities after excluding fewer than 10 attributable episodes due to unstable small samples

Unadjusted PAC Rates:		**Risk-Standardized PAC Rates (RSPR):**	
Median (IQR):	50% (44%, 55.6%)	Median (IQR):	50% (44%, 55.6%)
Range:	31.6% - 80%	Range:	31.6% - 80%

- Descriptive data on the patient, facility and payer are not provided. The developer provides "Overview" and "Drill Down" PAC rates in the spreadsheet demonstrating gap, though PAC rates for individual complications are not provided.
- The developer does not provide data on disparities.

Questions for the Committee:

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

○ *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

○ *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

○ *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Comments on Overview***:
   - This is a composite outcome measure. Potentially avoidable complications (PACs) are classified as both type I (directly related; e.g., bleeding) or indirectly related (type 2; e.g., sepsis, phlebitis, DVT).

1a. ***Committee's Comments on Evidence to Support Measure Focus:***
   - Approximately 40% of PCIs had a potentially avoidable complication (PAC)

1b. ***Committee's Comments on Performance Gap:***
   - Unadjusted PAC Rates:
     - Median (IQR): 50% (44%, 55.6%)
     - Range: 31.6% - 80%
     - Risk-Standardized PAC Rates (RSPR):
     - Median (IQR): 50% (44%, 55.6%)
     - Range: 31.6% - 80%
   - No disparities information is provided. I am unaware of disparities. I don't think that it should be identified as disparities sensitive.

1c. ***Committee's Comments on Composite Performance Measure:***
   - The stated reason that this is an any-or-none measure is that many of the complications are quite rare and enormous sample sizes would be required to be significant. The components are not weighted.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure assesses the rate of patients with 1 or more PAC(s) during index episodes. This new risk adjusted outcomes measure is specified for use at the individual clinician, group/practice, team, facility & integrated delivery system levels of analyses.
- The measure exclusively uses electronically available administrative claims data to calculation the measure score, and for this measure, better care equals lower scores.
- The developer describes non-patient-related PACs as controllable by facility processes without further analysis as other influencers that may contribute to PAC rates beyond the patient and facility (e.g., payer, access, suppliers, etc.).
- Patient- and claims-based exclusions are provided to promote the availability and consistency of claims data capture, including payer enrollment requirements, cost outliers of < 1% or > 99%, and claims not relevant to PCI.
- Developers provide a robust data set of administrative claims codes for PCI & PAC (PCI- & Patient Safety Failure-related) triggers, describe a 30-day look-back period and a 90-day look-forward time window.
- A calculation algorithm is provided, as well as ICD-9 & ICD-10 codes, though ICD-10 descriptions & an ICD-9 to ICD-10 crosswalk methodology are not provided.

- A conceptual risk model and statistical method using logistic regression model for determining the probability of a patient incurring a PAC are provided. After adjusting for patient-related factors, the developers state the remaining PAC variance is due to factors potentially controlled by the facility during and after hospitalization. "Predicted" coefficients from the risk adjustment models are summed to give predicted probabilities of PAC occurrence.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- The developer tested reliability at the performance measure score, and used a beta-binomial model and a signal-to-noise analysis, which is appropriate for this type of measure, to differentiate the true difference between measured entities (the signal) to random measurement error (the noise). A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in between *facility* performance. A value of 0.7 is often regarded as a minimum acceptable reliability value, and the developer also states values above 0.9 are considered sufficient to see differences between pairs of physicians.
- The measure is specified for patients ≥ 18 years that underwent PCMDFR, though the testing sample includes patients 18 through 64 years.
- The measure is specified for use with individual clinician, group/practice, team, facility & integrated delivery system levels of analyses, though testing is provided for facilities. NQF's measure evaluation criterion requires testing for all measure specification levels.
- Facilities with < 10 PCI episodes were excluded from reliability testing, though the measure is specified for patient without episode restrictions. A sample of 565 facilities was initially included in the data set, though facilities with less than 10 PCI episodes were excluded, allowing for 41 remaining facilities. There were 5,898 episodes of PCI with a mean age of 55.6 (18-64 years) and 31% female in the testing analysis exclusively using administrative claims data.
- A patient may have more than one condition-specific concurrent episode with claims applied to both episodes. If an inpatient claim corresponds to a procedure index episode and to a condition index episode, the claim would be assigned to the procedure index episode, rather than the condition index episode (e.g., for a claim that corresponds to both index episodes of CAD & CABG, the claim would be assigned to CABG).
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- Reliability results are provided in the table below, as well as in great detail in the accompanied spreadsheet with median (IQR) results demonstrating reliability of 0.51 (0.26, 0.62) for ≥ 10 PCIs per facility, and 0.74 (0.70, 0.83) ≥ 175 PCIs per facility. For reliability analysis, facilities were restricted to the minimum of 10 PCI episodes, though all episodes were included in the risk model.
- The developer states, "Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset will apply to another." The developer also states that very high sample sizes are to achieve any meaningful and reliable comparisons.
- The developer provides a supplementary fact sheet (available for review on SharePoint) requiring a minimum of 185 index episodes for absolute reliability, and a minimum of 175 index episodes for median reliability of > 0.7.
- The developer provides an additional supplementary fact sheet related to PAC development & testing (available for review on SharePoint).

Reliability Scores	Minimum # Episodes Per Facility	
	>=10	>=175

# of Facilities (%)	41 (100)	8 (20)
Median (IQR)	0.51 (0.26,0.62)	0.74 (0.70,0.83)
Range	0.11-0.85	0.69-0.85

The table provides a summary reliability scores minimum sample size thresholds. Complete results are provided in the workbook entitled, NQF_PCI_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab.

Questions for the Committee:

- *Reliability testing was attempted only for those facilities with at least 10 episodes. Can differences in performance be identified for facilities with fewer than 10 or 175 episodes? For patients ≥ 65 years?*
- *Should the measure be specified to include only those facilities with at least 10 episodes?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- Because this is an outcome measure, the rationale that is presented for subcriterion 1a does not necessarily have to address all of the variables used to calculate the measure.
- The measure uses a statistical risk model with 170 risk factors and episode-specific subtypes/severity markers including age, gender and clinical comorbidities, on at least 10 claims to determine "stable" covariates and assess comorbidity or procedure impact on the PAC. All covariates must be present prior to an episode trigger. No formal covariate analysis was conducted to select risk factors beyond the present on a minimum of 10 claims threshold. The developer describes the heterogeneity of the provided data sets as crucial to ensure measure validity, and the accuracy and completeness of the data sets.
- The developers did not provide disparities data, an exploration of a conceptual relation to SDS, or SDS factors in the risk model.

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Are these variables available and generally accessible for the measured patient population?*

2b2. Validity testing

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.

- The developer conducted systematic assessment of face validity for the performance measure score for validity testing in numerous ways, including the use of monthly multi-specialty clinical working groups, and other tests of face validity, along with focus groups, face validity comparisons of the measure to other national accountability measures, as well as additional literature for the measure & PAC development process.
- No empiric results are provided for the face validity tests described above.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer describes patient- (demographic, enrollment or low/high claims cost) and claims-based (due to missing or non-relevant data) exclusions for the measure.
- A significant number of patients were eliminated from the measure due to exclusion criteria, including 5,898 of 10,177 (58.0%) PCI episodes (in 3,258,706 unique beneficiaries) and 41 of 565 (7.1%) facilities.

Questions for the Committee:

- *Are high cost outliers (> 99%) exclusions an opportunity to identify PACs?*
- *Does the high number of exclusions restrict the measure use?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across facilities to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- The risk model (detailed in the accompanied spreadsheet) includes 170 factors and subtypes including age, gender, 12-month enrollment markers, co-morbidities, and episode severity markers.
- No SDS factors beyond age and gender was included in the risk-adjustment approach. Beyond noting that race was not available for analysis, no description of the of the conceptual relationships between patient sociodemographic factors, patient clinical factors, quality of care, and the outcomes (PAC rates) were provided, nor do they discuss the availability of SDS variables, beyond stating that "race" as an SDS variable was not available for analysis.
- Logistic regression was used to model the probability of at least one PAC during an episode. The reasoning for no additional modeling performed is described.
- The performance of the model was determined with a split sample method by estimating the model coefficients using a development dataset (80% of the sample) and applying those coefficients to a validation dataset (20% of the sample). C-statistics for the development and validation samples with c-statistic results of 0.803 and 0.792, respectively. C-statistics measures the extent of a statistical model to discriminate between a patient with and without PAC, with an ability to predict if a PAC is or is not present about 75% of the time. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong.
- Both Hosmer-Lemeshow Goodness-of-Fit statistics and risk-decide plots were provided to indicate model fit. Results from the Hosmer-Lemeshow test suggest that the fit is not good; however, this test is sensitive to the number of groupings and sample sizes. Results from the risk decile plot indicate that the predicted PAC rates are similar to the observed PAC rates across all deciles of risk. The developer states the model demonstrates strong predictive power.

Questions for the Committee:

- *Is the Committee aware of conceptual relationship(s) between additional patient-level SDS factors, patient clinical factors, quality of care, and PACs (other than gender and age)? If so, what data might be available to allow an empirical analysis of these relationships?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*

2b5. Meaningful difference:

- The developer presents PAC rates across facilities and also facilities adjusting for differences in patient severity in a ratio of observed to expected attributable episodes to PACS accounting for patient severity, and calculates estimates from the risk model, for risk-standardized PAC rates for the facility.

Summary of Unadjusted and Adjusted Performance Scores Across Facilities:

PAC Rates	Minimum # Episodes Per Facility	
	>=10	>=175
Unadjusted		
Median (IQR)	50% (44%, 56%)	47% (44%, 52%)
Range	32% - 80%	43% - 58%
Adjusted (RSPR)*		
Median (IQR)	49% (44%, 52%)	49% (46%, 52%)
Range	24% - 61%	45% - 56%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_PCI_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- As there is only one data source used for measure calculation (administrative claims), comparability of data sources or methods is not applicable.
-

2b7. Missing Data

- No formal analysis of missing data is provided. As the measure assesses the rate of patients with PACs, rather than the rate of PACs per index episode, the total number of PACs is not included in the PAC rate.
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- The developers state the under-coding of claims is unavoidable in the current DRG payment structure which could lead to under capture or missing PACs.

2d. Empirical Analysis to Support Composite Construction

- As an "any or none" composite, the individual complications are considered measurable components of the composite. Frequency and distribution statistics are provided in the PAC Overview and PAC Drill Down tabs in the measure spreadsheet, which detail PAC types and subtypes. The identification of individual PACs are not provided (e.g., sepsis, unattended falls, DVT).
- PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.
- The developer states that no formal analysis was performed on missing data. For details, see 2b7 above.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- All of the data elements appear to be clearly defined.

2a2.: ***Committee's Comments on Reliability-Testing:***

- Data presented by the stewards suggest that adequate reliability is achieved with >=175 episodes per facility. However, they issue a caveat that each data set has its own characteristics and that the number of cases required to discriminate between two providers cannot be specified a priori
  - Questions for the Committee:
    - Are all the data elements clearly defined? Are all appropriate codes included? yes
    - Is the logic or calculation algorithm clear? yes
    - Is it likely this measur0e can be consistently implemented? yes, but see above for a caveat about sample sizes
- Testing was only done at the facility level.

2b1.: ***Committee's Comments on Validity-Specifications***:

- No threats to validity.

2b2.: ***Committee's Comments on Validity-Testing:***

- The measure was developed with input from a large number of clinicians. This creates face validity.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Risk adjustment is by age, gender, recent enrollment and clinical co-morbidities.
- Text suggests that results may be sensitive to coding practices within an institution and the stewards cite references that document under-reporting. This suggests that absolute rates may not be comparable across institutions.

2d.: ***Committee's Comments on Composite Performance Measure:***

- The composite does fit the quality construct and appears to add value.

Criterion 3. Feasibility

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

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than the person obtaining the original information (e.g., DRG, ICD-9 codes on claims).
- The developer provides an excel spreadsheet attachment including diagnoses, visits, hospitalizations, post-acute facility stays, procedures, laboratory tests and procedures/surgeries, for CAD & PAC triggers, and describe the time window for measuring PAC triggers as a 30 day look back and 90 days after undergoing PCI, as well as a decision tree for measure calculation and implementation.
- The measure specifications, metadata and calculation algorithms are available free of charge on the developer's website. Limited analytics are planned at no cost to the end user.
- This is not an eMeasure.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- The measure is entirely claims based, so all data are routinely collected. However, the developers present data that indicate that PaCs are under-reported.
  - Questions for the Committee:
    - Are the required data elements routinely generated and used during care delivery? A qualified "yes". The stewards write that "when the claims datasets are organized in the way we specify..." the results are highly reliable. This suggests to me that there are cases in which the feasibility of producing a reliable analysis is a problem.
    - Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
    - Is the data collection strategy ready to be put into operational use

Criterion 4: Usability and Use

4. Usability and 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.

- This is a newly developed claims measure is current used in accountability programs for payers, states, and planned for public reporting, professional certification or recognition programs, and external quality improvement for benchmarking purposes.
- The developer states that PAC measures provide a foundation for the relationship between healthcare quality and cost and assist in the exploration of practice reengineering and alternative payment models, act as indicators of potential harm, and is spurring the development of private-based analytics software for further outcomes exploration. No public improvement rates are available due to recent implementation and variation in PAC definitions have also modified, though the provided PROMETHEUS data suggest wide variation in performance and improvement opportunities.
- The developer found no noted unintended consequences, though they state the measure is intended for transparency and QI activities only. They also state the under-coding of claims is unavoidable in the current DRG payment structure could be an unintended consequences of the measure, and payers calculating the measures even with inadequate sample sizes and using the results to penalize providers could lead to invalid provider comparisons.
- If the measure was theoretically to be used for accountability purposes to "ding" the measured entity as defined in the level of analysis, further exploration of PAC antecedents and the measured entity is warranted, especially with lower volume PCI facilities. In the original testing sample of 565 facilities, when facilities with fewer than 10 PCI episodes were eliminated from analysis due to less reliability estimates with small numbers, 41 remained for analysis.

***Questions for the Committee**:*

- *Is the measure publicly reported?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Should PAC measures also include the clinician: group in the analysis or include population-level only entities?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use***

- Not currently publicly reported.
  - Current use:
    - Payment Programs
    - Blue Cross Blue Shield of North Carolina
    - Blue Cross Blue Shield of New Jersey
    - Pennsylvania Employee Benefits Trust Fund
  - Quality Improvement (Internal to the specific organization)
    - Blue Cross Blue Shield of North Carolina
  - Planned use
    - Public Reporting
    - Professional Certification or Recognition Program
    - Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

Criterion 5: Related and Competing Measures

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

0141 : Patient Fall Rate
0202 : Falls with injury
0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0695 : Hospital 30-Day Risk-Standardized Readmission Rates following Percutaneous Coronary Intervention (PCI)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the

30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

Pre-meeting public and member comments

Comment by: Ashish R. Trivedi, Pharm.D.
Organization: SPI-Lilly
Comment#5115: Lilly is supportive of this measure as it focuses on reducing risk for potentially avoidable (eg, via improvement in quality of treatment and care) recurrent major adverse cardiovascular events (MACE), at no expense of increased safety events.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2751
Measure Title: Proportion of Patients undergoing an Angioplasty Procedure (Percutaneous Coronary Intervention - PCI) that have a Potentially Avoidable Complication (during the episode time window)
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: 6/30/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.

- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☒ Health outcome: Potentially Avoidable Complications

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

_____

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

This measure addresses potentially avoidable complications (PACs) in patients undergoing PCI. PACs could be both directly or indirectly related to healthcare services provided (or not provided) for a condition or procedure. When complications occur within PCI, the impact could be catastrophic for the patient. Peri-procedural bleeding which carries a high mortality rate, is more likely to occur in less competent hands or the need for emergency CABG is higher in low volume centers (Crudu 2011) (Harold 2013). The teams' experience, and the volume of PCIs in the center, plays a role in reducing complications like the need for emergency CABG, or periprocedural bleeding. Developed in 1997 by the American College of Cardiology, the National Cardiovascular Data Registry (NCDR 2015) is the most comprehensive database of post-PCI patients and a method by which hospitals are compared for quality of care and outcomes (NCDR 2015). This gives centers performing PCI's an opportunity to assess their efficacy and give an opportunity to improve quality at the provider, hospital, and/or health care system level. Some studies have shown that actual rate of complications post-PCI may be higher than the rate reported in NCDR. Access site related complications occurred 13% more than what was reported in NCDR and were associated with a greater than fourfold increase in in-hospital mortality (Crudu 2011). Additionally, a study

from the Mayo Clinic showed that nearly 1 in 10 patients undergoing PCI were readmitted within 30 days and were associated with a higher risk of 1-year mortality (Khawaja 2012).

Many of these adverse events are aggregated together in the PAC measure to study the overall rate of avoidable complications in the PCI population.

Adult patient undergoing a PCI procedure admitted to hospital or an outpatient facility

↓

Hospital/physician fails to carry out safe practices (error in commission/omission)

↓

Patient suffers complication stemming from hospital/physician potentially avoidable error

↓

Patient remains in hospital for treatment of PAC

OR

Patient readmitted to hospital with 1+ Potentially avoidable complication

Well-managed patients undergoing a PCI procedure should rarely incur a potentially avoidable complication such as bleeding, AMI, stroke, and readmissions after PCI should occur only in the rarest of circumstances. The enclosed workbook entitled NQF_PCI_all_codes_risk_adjustment 06.30.15.xls lists the types of PACs and their frequency as calculated in a large regional database (see tab PAC overview). Over 47% of PCI episodes had a PAC. Of these, 37.5% were incurred for direct complications of PCI, such as hypotension, syncope, complication of stents, shock and cardiac arrest (see tab PAC Drill Down Graph). Although the preventable readmissions in the PCI population were low, at only 4.2%; approximately 21.6% of patients with PCI procedures had PACs related to patient centered care failures such as poor control of diabetes, respiratory failure, and phlebitis and deep vein thrombosis, many of them being managed in an outpatient setting in physician offices. As a result 38.7% of episodes had a PAC indicator on their professional claims.

While PACs may not be eliminated completely, identifying the magnitude of PACs and knowledge of the cause for the most frequent or the most expensive PACs could place an emphasis in reducing them and as a consequence improving patient outcomes. The ability to clearly identify the type and frequency of each PAC creates a highly actionable measure for all providers that are managing or co-managing the patient, as well as for the health plan with whom the patient is a member.

References:

1) Crudu V, Blankenship J, Berger P, Scott T, Skelding K. “Complications related to access site after percutaneous coronary interventions: are the adverse events underreported?” *Catheter Cardiovasc Interv* 77 (2011):643–7.

2) Harold, John G., et al. "ACCF/AHA/SCAI 2013 Update of the Clinical Competence Statement on Coronary Artery Interventional Procedures." *Catheterization and Cardiovascular Interventions Cathet. Cardiovasc. Intervent.* 82.2 (2013). Web.

3) "CATH PCI Registry - Institutional Outcomes Report 2014Q3." *National Cardiovascular Data Registry*. American College of Cardiology Foundation, 29 Jan. 2015. Web. <http%3A%2F%2Fcvquality.acc.org%2F~%2Fmedia%2FQII%2FNCDR%2FSample%2520Reports%2FCathPCI_Registry_2014Q3_Sample_Report.ashx>.

4) Khawaja, FJ, et al. "Factors associated with 30-day readmission rates after percutaneous coronary intervention". *Arch Intern Med*. (2012): 172:112-117.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

Rationale: Better processes of care create an atmosphere of proactive management, consistency in care and standardized care patterns (Shekelle 2013) (Fenter 2006). Patient education and adopting safe practices significantly reduces occurrence of potentially avoidable complications (PACs) in all settings (Klein 2011) (Wachter 2013) (Berwick 2006) (Kovner 2011) (Farley 2013). It is known that by holding providers accountable for occurrence and costs of PACs, an built-in warranty is created around care of the index condition (de Brantes 2009).

Percutaneous coronary intervention (PCI) is the preferred treatment strategy for ST-elevation myocardial infarction (STEMI) (Langabeer 2013). National quality improvement initiatives have focused on patient education and awareness of acute MI symptoms in order to reduce door-to-balloon times for primary PCI and improving outcomes in acute MI patients (Jollis 2007) (Jacobs 2007).

However, major PCI-related complications are not uncommon such as death, MI, emergency CABG surgery, and stroke, commonly denoted as MACCE (major adverse cardiovascular and cerebrovascular events). Other complications are vascular, bleeding and contrast nephropathy (Harold 2013). Low volume centers and low volume operators have a higher rate of complications for PCI, therefore the ACC/AHA/SCAI have adopted standards for minimally acceptable patient procedure volumes to be performed by cardiologists and hospitals (75 cases per annum per operator) (Harold 2013). The ACCF/AHA/SCAI in its 2013 Update of the Clinical Competence Statement on Coronary Artery Interventional Procedures has created guidelines to facilitate the attainment of optimal patient outcomes such as selection of clinically appropriate patients for interventional procedures and performing these procedures at a requisite level of proficiency and competency (Dehmer 2014). Opportunities for improvement exist in PCI by using the available tools (for example SYNTAX, ACC/AHA score and SCAI score) to optimize clinical decision-making to reduce complications and improve outcomes (Harold 2013).

In addition, to the direct complications related to the PCI procedure itself, there are several patient safety failures that may occur. There are a wide variety of ways to reduce these potentially avoidable complications or PACs. A few examples of better processes of care leading to reduced PACs are given below:

1. Good surgical technique reduces bleeding, perforation and complications directly from the procedure
2. Frequent hand-washing reduce hospital acquired infections (WHO 2007)
3. Carefully implemented protocols lead to reduced line sepsis (Pronovost 2010)
4. Aspirin on Discharge prevents repeat AMIs (Hall 2014)
5. Discharge planning and good follow-up prevents unnecessary ER visits and readmissions (Weaver 2013)
6. DVT prophylaxis in patients on bed rest avoids pulmonary embolism (Shekelle 2013)

PAC measures in the setting of PCI look at all-cause harms, such as the ones highlighted above, arising from poor management of a patient undergoing an angioplasty procedure.

References:

1) Shekelle, Paul G., et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

2) Fenter TC, et al. "The cost of treating the 10 most prevalent diseases in men 50 years of age or older." *Am J Manag Care* 12.4 (2006): S90-S98.

3) S. Klein and D. McCarthy, OSF HealthCare: "Promoting Patient Safety Through Education and Staff Engagement." *The Commonwealth Fund* 1475.7 (March 2011).

4) Wachter, Robert M., Peter Pronovost, and Paul Shekelle. "Strategies to Improve Patient Safety: The Evidence Base Matures." *Annals of Internal Medicine* 158.5_Part_1 (2013): 350-52. Web.

5) Berwick, D.M. D. R. Calkins, C. J. McCannon et al., "The 100,000 Lives Campaign: Setting a Goal and a Deadline for Improving Health Care Quality," *Journal of the American Medical Association*,295.3 (2006):324–27

6) Kovner, Anthony R., James Knickman, and Steven Jonas. "Chapter 11." *Jonas & Kovner's Health Care Delivery in the United States*. 10th ed. New York: Springer Pub., 2011. Print.

7) Farley, TG, et al. "The Top Patient Safety Strategies That Can Be Encouraged for Adoption Now." *Annals of Internal Medicine* 158.5_Part_2 (2013): 365-69. Web.

7) de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? *Health Affairs (Millwood)* 28 (2009): w678-w687.

8) Langabeer, J. R., T. D. Henry, D. J. Kereiakes, J. Dellifraine, J. Emert, Z. Wang, L. Stuart, R. King, W. Segrest, P. Moyer, and J. G. Jollis. "Growth in Percutaneous Coronary Intervention Capacity Relative to Population and Disease Prevalence." *Journal of the American Heart Association* 2.6 (2013). Web.

9) Jollis, J.G., M.L. Roettig, A.O. Aluko, *et al.* Implementation of a statewide system for coronary reperfusion for ST-segment elevation myocardial infarction. *JAMA*, 298 (2007), pp. 2371–2380

10) Jacobs, A.K., E.M. Antman, D.P. Faxon, T. Gregory, P. Solis. "Development of systems of care for ST-elevation myocardial infarction patients: executive summary." *Circulation* 116 (2007), pp. 217–230

11) Harold, John G., et al. "ACCF/AHA/SCAI 2013 Update of the Clinical Competence Statement on Coronary Artery Interventional Procedures." *Catheterization and Cardiovascular Interventions Cathet. Cardiovasc. Intervent.* 82.2 (2013). Web.

12) Dehmer, G. J., J. C. Blankenship, M. Cilingiroglu, J. G. Dwyer, D. N. Feldman, T. J. Gardner, C. L. Grines, and M. Singh. "SCAI/ACC/AHA Expert Consensus Document: 2014 Update on Percutaneous Coronary Intervention Without On-Site Surgical Backup." *Circulation* 129.24 (2014): 2610-626. Web.

13) The Joint Commission, Aide Joint Commission International, and WHO. "Improved Hand Hygiene to Prevent Health Care-Associated Infections." World Health Organization. WHO Collaborating Centre for Patient Safety Solutions (May 2007). Web.

14) Pronovost, P.J., G. A. Goeschel, E. Colantuoni et al., "Sustaining Reductions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study," *BMJ* 340:c309 (Feb. 4, 2010): Web.

15) Hall, Hurst M., MD, James A. De Lemos, MD, and Jonathan R. Enriquez, MD. "Contemporary Patterns of Discharge Aspirin Dosing after Acute Myocardial Infarction in the United States: Results from the National Cardiovascular Data Reg." Circulation: Cardiovascular Quality and Outcomes. American Heart Association Journals, 12 Aug. 2014. Web.

16) Weaver, Sallie J., Lisa H. Lubomksi, Renee F. Wilson, Elizabeth R. Pfoh, Kathryn A. Martinez, and Sydney M. Dy. "Promoting a Culture of Safety as a Patient Safety Strategy." Annals of Internal Medicine 158.5_Part_2 (2013): 369-75. Web.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

_____

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

2751_PCI_Evidence_Attachment_HCI3.docx

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., the benefits or improvements in quality envisioned by use of this measure)*

Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some

provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 40% of its plan members with hypertension incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow- up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.

4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

6) BCBSNC: Blue Cross Blue Shield of North Carolina:
https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5. http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/

11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.

14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

The data included two years of administrative claims covering the period April 1, 2012 through December 17, 2014. There were a total 5,898 episodes of PCI.

Because facilities with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. After this exclusion 41 (out of 565) facilities remained. Performance scores for these facilities are summarized in the following table:

Unadjusted PAC Rates:

Median (IQR):	50% (44%, 55.6%)
Range:	31.6% - 80%

Risk-Standardized PAC Rates (RSPR):

Median (IQR):	48.5% (43.7%, 52.2%)
Range:	24.2% - 61.5%

Please refer to the NQF_PCI_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

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.

Despite an aging population in the US, current trends point towards a decrease in hospitalizations related to coronary artery disease events. This applies to both a decrease in acute myocardial infarcts, due to decrease in transmural AMI, as well as decreased

hospitalization rates in patients undergoing revascularization procedures (Nallamothu 2007). In fact in this study, rates of hospitalizations for coronary revascularizations dropped from 382 to 358 per 100,000 between 2002 and 2005. Interestingly, the decrease was primarily due to a decrease in rates of CABG procedures from 121 to 94 per 100,000. Numbers of PCIs, however, continued to increase from 264 to 267 per 100,000. Interestingly, with better availability of drug-eluting stents, and better techniques in PCI there has been dramatically lower incidence of restenosis rates following PCI and therefore decease in re-hospitalizations. However, high degree of variability in PCI related complications across different datasets (de Brantes 2011) suggest that short-term complications should be measured and reported.

References:

1) Nallamothu BK, Young J, Gurm HS, et.al. "Recent Trends in Hospital Utilization for Acute Myocardial Infarction and Coronary Revascularization in the United States" Am J Cardiol 99 (2007): 749-753.

2) de Brantes F, Rastogi A, and Sorensen CM. "Episode of Care Analysis Reveals Sources of Variation in Costs." Am J Manag Care 17.10 (2011): e383-e392.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Not applicable

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Frequently performed procedure, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
PCI is the central therapy for patients with symptomatic coronary artery disease, particularly acute myocardial infarction (Roe 2010). Over 1 million Coronary revascularizations are performed annually, making it the most common major medical procedure in the US (Epstein 2008). It is also the most expensive procedure, with Medicare inpatient costs for revascularization procedures exceeding 3.2 billion annually in 2006 (Epstein 2008). The indications for PCI have been expanding, and technology including the availability of drug eluting stents has contributed to increased use of PCI (Epstein 2008). Longitudinal data obtained for 2003–2011 from the American Hospital Association, the U.S. Census, and the Centers for Disease Control and Prevention (CDC) showed centers performing PCI have grown 1.5 times faster than population growth in the same period (Langabeer 2013).

A study by de Brantes et al analyzed variability in episode costs in patients undergoing PCI and demonstrated that there was significant variation in episode costs within health plan data, and PACs contributed considerably to that variation (de Brantes 2011). This suggests that there is room for improvement. Another large study of over 400,000 patients across 1499 hospitals, analyzing 2005 Medicare FFS claims data, showed that approximately 1 in 7 Medicare patients undergoing PCI are readmitted within 30 days, and readmission rates vary across hospitals reflecting variations in quality of care (Curtis 2009). Readmissions are not only associated with higher revascularization rates (up to 25%), they also have a significantly higher mortality rate (Curtis 2009).

Major PCI-related complications include death, MI, emergency CABG surgery, and stroke, commonly denoted as MACCE (major adverse cardiovascular and cerebrovascular events). Other complications are vascular, bleeding and contrast nephropathy. Procedural success is defined as angiographic success without in-hospital major complications such as death, myocardial infarction (MI), stroke, and emergency coronary artery bypass graft (CABG) surgery (Harold 2013).

With the rise in use of PCI procedures, and the significant number of avoidable complications (Crudu 2011) (Mukherjee 2005) (Normand 2008) (Patel 2009) (Dehmer 2014) that are under the provider's control, PAC measures in the PCI population are a high priority aspect of health care. The PAC measures look for all-cause harms in patients receiving angioplasty.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1) Roe MT, Messenger JC, Weintraub WS, et al. "Treatments, trends, and outcomes of acute myocardial infarction and percutaneous coronary intervention." J Am Coll Cardiol 56 (2010): 254 – 63.

2) Epstein, Andrew J, et al. "Coronary Revascularization Trends in the United States, 2001-2008." JAMA 305.17 (2011): 1769. Web.

3) Langabeer, J. R., T. D. Henry, D. J. Kereiakes, J. Dellifraine, J. Emert, Z. Wang, L. Stuart, R. King, W. Segrest, P. Moyer, and J. G. Jollis. "Growth in Percutaneous Coronary Intervention Capacity Relative to Population and Disease Prevalence." Journal of the American Heart Association 2.6 (2013). Web.

4) de Brantes F, Rastogi A, and Sorensen CM. "Episode of Care Analysis Reveals Sources of Variation in Costs." Am J Manag Care 17.10 (2011): e383-e392.

5) Curtis, Jeptha P., Geoffrey Schreiner, Yongfei Wang, Jersey Chen, John A. Spertus, John S. Rumsfeld, Ralph G. Brindis, and Harlan M. Krumholz. "All-Cause Readmission and Repeat Revascularization After Percutaneous Coronary Intervention in a Cohort of Medicare Patients." Journal of the American College of Cardiology 54.10 (2009): 903-07. Web.

6) Harold, John G., et al. "ACCF/AHA/SCAI 2013 Update of the Clinical Competence Statement on Coronary Artery Interventional Procedures." Catheterization and Cardiovascular Interventions Cathet. Cardiovasc. Intervent. 82.2 (2013). Web.

7) Crudu V, Blankenship J, Berger P, Scott T, Skelding K. "Complications related to access site after percutaneous coronary interventions: are the adverse events underreported?" Catheter Cardiovasc Interv 77 (2011):643–7.

8) Mukherjee, D., R. M. Wainess, et al. "Variation in outcomes after percutaneous coronary intervention in the United States and predictors of periprocedural mortality." Cardiology 1033 (2005): 143-7.

9) Normand, S. L. "PERCUTANEOUS CORONARY INTERVENTION IN THE COMMONWEALTH OF MASSACHUSETTS Fiscal Year 2006 Report October 1, 2005 – September 30, 2006." Department of Health Care Policy (2008): n. pag. Mass DAC. Harvard Medical School, Feb. 2008. Web. <http://www.massdac.org/wp-content/uploads/PCI-FY2006-Update.pdf>.

10) Patel, M. R., G.J. Dehmer, et al. "ACCF/SCAI/STS/AATS/AHA/ASNC 2009 appropriateness criteria for coronary revascularization, J Am Coll Cardiol 53 (2009): 530–553.

11) Dehmer, G. J., J. C. Blankenship, M. Cilingiroglu, J. G. Dwyer, D. N. Feldman, T. J. Gardner, C. L. Grines, and M. Singh. "SCAI/ACC/AHA Expert Consensus Document: 2014 Update on Percutaneous Coronary Intervention Without On-Site Surgical Backup." Circulation 129.24 (2014): 2610-626. Web.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable

1d. Composite Quality Construct and Rationale

1d.1. **A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.

- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
    - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient); or
    - any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient).

1d.1. Please identify the composite measure construction: any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

1d.2. Describe the quality construct, including:

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

We classify PACs into two types: Type 1 PACs are directly related to the index condition and are often controlled by the servicing provider; Type 2 PACs, on the other hand result from patient safety failures and could be reduced by better systems and better processes in care. Both types of PACs could occur in any setting and so could be identified through any type of claims coming in the administrative dataset, including in-patient, out-patient, or professional claims. PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1.

The PAC measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs), Hospital Inpatient Quality Reporting measures, Avoidable Readmissions, AHRQ defined patient safety indicators (PSIs), NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates.

All defined PACs, irrespective of their type, or site of occurrence, are aggregated to create an overall comprehensive, composite measure. They all have equal weighting, since they are measured simply by the frequency of their occurrence.The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

We classify PACs into two types: Type 1 PACs are directly related to the index condition and are often controlled by the servicing provider; Type 2 PACs, on the other hand result from patient safety failures and could be reduced by better systems and better processes in care. Both types of PACs could occur in any setting and so could be identified through any type of claims coming in the administrative dataset, including in-patient, out-patient, or professional claims. PACs may occur any time during the episode time window. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence.

The PAC measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs), Hospital Inpatient Quality Reporting measures, Avoidable Readmissions, AHRQ defined patient safety indicators (PSIs), NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates.

All defined PACs, irrespective of their type, or site of occurrence, are aggregated to create an overall comprehensive, composite measure. They all have equal weighting, since they are measured simply by the frequency of their occurrence.

1d.3. Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.

Each individual PAC, when measured in isolation, provides a very limited picture of the performance of the provider(s) who are managing or co-managing the care of the patient. However, looking at all the PACs that may occur individually or concurrently in a patient with a given episode provides a comprehensive picture of the care received by the patient for that particular condition or illness.

Additionally, the frequency of occurrence of individual PACs may be so low that it may require very high sample sizes from individual

providers to achieve any meaningful and reliable comparisons. But aggregating all the PACs into a single quality metric creates meaningful scores that can be compared across providers even with relatively smaller sample sizes.

Additionally, a comprehensive measure is easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

1d.4. **Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**
In constructing the comprehensive composite PAC measure, each component PAC, as clinically defined by the subject matter experts, was given the same weight so that arbitrary weights may not bias the results. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence. As such, the patient is the ultimate unit of measurement and if the patient incurred any PAC during the episode, then that counts against the numerator.

Since the emphasis of the PAC measure was to simply identify the occurrence of PACs in any setting, aggregation of the PAC counts to create a comprehensive quality score with equal weights has been met with overall support from the clinical working groups as well as from the implementation sites.

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 subcriteria 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 : Percutaneous Coronary Intervention (PCI)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Healthcare Associated Infections, Safety : Medication Safety, Safety : Readmissions, Safety : Venous Thromboembolism

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.)*
http://www.hci3.org/ecr_descriptions/ecr_description.php?version=5.2.006&name=PCI&submit=Submit

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_PCI_all_codes_risk_adjustment_06.30.15-635719835998602641.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable

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)*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of patients who underwent a percutaneous coronary intervention (PCI) procedure, are followed for at least 90-days, and have one or more potentially avoidable complications (PACs) during the episode time window.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The time window includes a 30-day look-back period and a 90-day look-forward period from the PCI trigger claim.

S.6. **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, 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.
Patients that have triggered a PCI episode, are followed for at least 90-days, and are identified as having services for potentially avoidable complications (PACs). PACs may occur during the index stay or during the 90-day post discharge period. The enclosed excel workbook entitled NQF_PCI_all_codes_risk_adjustment_06.30.15 gives the detailed codes for PACs in the tabs entitled PACs I-9 and PACs I-10.

Services for PACs are identified as follows:
a. Any Index stay that has a PAC diagnosis code in any position except in the PRIMARY (principal) position is considered as having a potentially avoidable complication
b. Any readmission to an acute care facility 2 days or later after discharge but within 90-days post-discharge, that is relevant to PCI
c. Any admission to a post-acute care facility, that is relevant to PCI and has a PAC code in any position on the claim
d. Any other service (professional, outpatient facility, ancillary) that is relevant to PCI and has a PAC code in any position on the claim

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Adult patients aged 18 years and above who underwent an Angioplasty (percutaneous coronary intervention - PCI) procedure and are followed for at least 90-days

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Please refer to the enclosed excel workbook entitled
NQF_PCI_all_codes_risk_adjustment 06.30.15

The target population is identified using the following criteria:
1. Using administrative claims database, patients undergoing PCI are identified using one of the following criteria:
a. Patients with a procedure code of PCI in any position on an in-hospital stay claim with a qualifying diagnosis code relevant to the PCI procedure.
b. Patients with a procedure trigger code of PCI in any position on an outpatient facility claim with a qualifying diagnosis code relevant to the PCI procedure.
c. Patients having a professional service carrying a trigger code of PCI in any position.
The trigger codes for PCI and the qualifying diagnosis codes are provided in the tab called “Triggers I-9” or “Triggers I-10”.

2. The patient should have continuous enrollment for the entire time window with no more than 30 days as an enrollment gap, with the entity providing the data (so we can ensure that the database has captured most of the claims for the patient during the episode time window).
3. The patient should have a complete episode time window in the claims data – so the end date of the episode should not be past the database claims end date.

4. Patient should be at least 18 years of age
5. Patients that have a trigger code on a professional claim and have no associated facility bill are considered as having an orphan (incomplete) episode and are dropped from analysis.

Once the episode is triggered all relevant claims are assigned to the episode. Relevant claims could be inpatient facility claims, outpatient facility claims, professional services, laboratory services, imaging services, ancillary claims, home health, durable medical equipment as well as pharmacy claims across the entire continuum of care centered around the patient's episode of care. Relevant claims are identified as those that have a diagnosis code that matches the codes in the typical Dx codes tabs (Typical Dx I-9 or Typical Dx I-10), or in the PAC Dx codes tab (PACs I-9 or PACs I-10) AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant readmissions and relevant admissions to post-acute care facilities are also included in the denominator.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for more than 30 days during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the PCI measure if they are considered not relevant to PCI care.

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Denominator exclusions include exclusions of "patients" as well as "claims" not relevant to PCI care. Please refer to the enclosed excel workbook entitled (NQF_PCI_all_codes_risk_adjustment 06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
 a. If age is < 18 years
 b. If gender is missing
 c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
 d. If the episode time window extends beyond the dataset end date (this helps eliminate incomplete episodes).
 e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure based on the following criteria:
 a. If none of the diagnosis codes on the claim are on the list of relevant diagnosis codes (either typical Dx or PAC Dx) for PCI.
 b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for PCI.

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
None

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Conceptual Model
Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, the remaining variance in PACs are due to factors that could be controlled by all providers that are managing or co-managing the patient, both during and after hospitalization.

Statistical Method:
Logistic Regression model to determine the probability of a patient incurring a PAC
Demographic variables, comorbid conditions, as well as clinical severity indicators are fed as independent risk factors into the model. Risk Factors are collected historically. Subtype information is collected from the index claim and any look-back period, if relevant. Subtypes are clinical severity indicators suggesting severity of the episode itself, for example, diagnosis of unstable angina in a PCI patient. For each patient the "predicted" coefficients from the risk adjustment models are summed to give the predicted probabilities of the occurrence of a PAC.

Risk Factors :(Please refer to the enclosed excel workbook entitled (NQF_PCI_all_codes_risk_adjustment 06.30.15.xls). The risk factors along with their codes are listed in the tabs called "All Risk Factors I-9" and "All Risk Factors I-10" and also listed below:

AGE CONTINUOUS VARIABLE
GENDER FEMALE = 1 (MALE IS REFERENCE = 0)

Risk Factor #	Risk Factor Name
RF0101	Anoxic Brain Damage, persistent vegetative state
RF0102	Delirium, Meningitis, Encephalitis
RF0103	Previous Stroke, Paralysis
RF0104	Cerebral Palsy and Other Paralytic Syndromes
RF0105	Spinal Cord Disorders/Injuries
RF0106	Polyneuropathy
RF0107	Multiple Sclerosis
RF0108	Convulsions, Epilepsy
RF0109	Dementia
RF0110	Parkinson´s and Huntington´s Diseases
RF0111	Cerebrovascular Disease
RF0115	after care, rehabilitation
RF0201	visual loss, blindness, retinal tear, detachment
RF0301	ENT, Upper Respiratory Problems
RF0401	Respiratory Failure, O2, ventilator dependence
RF0402	Advanced COPD, Asthma
RF0403	Empyema, bronchiectasis, Pneumonias
RF0404	Aspiration Pneumonia, Laryngeal Problems
RF0406	TB, Pneumoconiosis, Aspergillosis
RF0407	Tobacco use, Lung disease due to External Fumes
RF0408	Other Lung Disease
RF0501	Previous Shock, Syncope, Vent Fibrillation
RF0503	Advanced CHF
RF0504	Cardiomyopathy, valve disorders
RF0505	Cardiac Arrhythmias, Heart Block
RF0506	Pacemaker, AICD
RF0507	Endocarditis, Other post surgical cardiac problems
RF0508	Other Cardiovascular Disease
RF0511	DVT, Pulm Embolism, Pulm Heart Disease
RF0512	Unstable Angina
RF0513	Hypotension, chronic, orthostatic
RF0514	Hyperlipidemia
RF0515	Intraaortic Balloon Pump
RF0516	ventricular assist device, ecmo, prolonged bypass
RF0517	Previous electrophysiology studies, cryoablation
RF0518	Recent AMI
RF0519	Previous PCI
RF0520	Previous CABG
RF0521	Previous Heart & Valve Surgery
RF0522	Previous aortic reconstruction
RF0523	Previos carotid endarterectomy

RF0524 Aortic and peripheral vascular disease
RF0525 Advanced Aortic and Vascular Disease
RF0601 GI Bleed
RF0602 Intestinal Obstruction/Perforation
RF0603 Acute Gastritis, Duodenitis
RF0604 Gastroduodenal Ulcer
RF0606 Intestinal Uro-genital Fistula
RF0607 Abdominal hernia w complications
RF0608 Vascular insufficiency of intestine
RF0609 Inflammatory Bowel Disease
RF0610 Irritable Bowel
RF0611 Diverticulitis, Meckel´s
RF0612 Digestive congenital anomalies
RF0613 Intestinal infection
RF0614 Esophageal Perforation, Hmg, Barretts, Compl Hiatal Hernia
RF0615 Abnormal weight loss
RF0616 Achalasia, Esophageal spasm, Stricture, Dysphagia
RF0617 GERD, Hiatal Hernia, Other Upper GI Disorders
RF0618 Previous Bariatric Surgery
RF0619 Hx of colon polyps, family Hx of colon cancer
RF0620 Enterostomy, GI devices, lap band
RF0701 Pancreatic Disease
RF0702 Perforation, fistula GB, bile duct, pancreas
RF0703 Gall stones, cholecystitis
RF0704 End-Stage Liver Disease
RF0705 Hepatitis, Cirrhosis, Other Hepatbiliary Disorders
RF0706 Recent Gall Bladder, Hepatobilary Surgery
RF0707 Acute Pancreatitis, pseudo cyst
RF0801 Bone/Joint/Muscle Infections/Necrosis
RF0802 Muscular Dystrophy
RF0803 Osteoporosis, ostetits deformans, pathological fracture
RF0804 Rheumatoid Arthritis and Inflammatory Connective Tissue Disease
RF0805 Gout and other crystal arthropathies
RF0806 Other arthropathies
RF0807 Osteoarthritis
RF0808 Joint Deformities
RF0809 Knee derangements
RF0810 Traumatic Dislocation Knee
RF0811 Dislocation Hip
RF0812 Synovitis, Ruture Tendon
RF0813 Status Knee Replacement
RF0814 Status Total Hip Replacement
RF0901 Decubitus Ulcer
RF0902 Skin and wound problems
RF1001 Diabetes, poor control
RF1002 Advanced diabetes
RF1003 diabetes
RF1101 Acute renal failure
RF1102 Dialysis Dependent
RF1103 Nephritis
RF1104 Chronic renal failure
RF1105 Urinary Tract Infections
RF1301 Endometriosis
RF1302 Fibroid uterus, benign tumors of female organs
RF1303 Pelvic Inflammatory disease
RF1304 Uterine prolapse, cystocele, vaginocele
RF1305 Female Harmonal Disorders

RF1306 Ovarian, Broad Ligament Disorders
RF1308 Other disorders of uterus, cervix
RF1309 Menopausal Disorders
RF1310 Menstrual Disorders
RF1401 Multiparity, multigravida
RF1402 Elderly Primi, other
RF1403 Poor obstetric history
RF1406 Cervical incompetence
RF1407 Abnormalities of uterus, female genital tract
RF1408 Hypertension, pre-eclampsia in Pregnancy
RF1409 Severe pre-eclampsia w HTN, Eclampsia
RF1410 Maternal, gestational diabetes, large for date
RF1411 Genital Herpes
RF1412 Infections of genitourinary tract, venereal disease in pregnancy
RF1413 Infectious Diseases in Mother
RF1414 Cardiovascular disease in Mother
RF1415 Mental Disorders in Mother
RF1416 Epilepsy in Mother
RF1417 Liver and biliary tract disorders in mother
RF1418 Kidney Disease in Mother
RF1419 Other Maternal conditions
RF1421 Cephalopelvic Disproportion due to maternal causes
RF1436 Peripartum Cardiomyopathy
RF1441 Previous Cesarean section
RF1450 Maternal Obesity, previous Bariatric Surgery
RF1454 Previous Rupture Uterus, Obstetrical Trauma
RF1458 Complicated Pregnancy Delivery
RF1460 Thrombophlebitis, DVT during Pregnancy
RF1461 Puerperal Sepsis, other major puerperal complications
RF1462 Obstetrical Embolism, Air, Amniotic Fluid, Pulm, Pyemic
RF1467 Tobacco Use in Mother
RF1601 Bleeding Disorders
RF1602 Severe Hematological Disorders
RF1603 Disorders of Immunity
RF1604 Nutritional and other Anemias
RF1605 Long-term use of anticoag, Aspirin
RF1701 Head and Neck Cancers
RF1702 Lung and Intrathoracic Cancers
RF1703 Neuroendocrine, Myeloproliferative Cancers
RF1704 Poorly differentiated, Secondary, Metastatic Cancers
RF1705 Other Tumors
RF1706 Acute Leukemia
RF1707 Cancer uterus, localized female organs
RF1708 Colorectal, Hepatobiliary and other GI cancers
RF1709 Breast, Prostate, Thyroid cancers
RF1710 Testicular Cancer and localized of male organs
RF1711 Cancer of Bladder and Urinary Tract
RF1712 Musculoskeletal Cancers
RF1801 Sepsis, MRSA, Opportunitistic infections
RF1901 Schizophrenia
RF1902 Major Depressive, Bipolar, and Paranoid Disorders
RF2001 Drug/Alcohol Psychosis
RF2002 Drug/Alcohol Dependence
RF2101 Drug Reactions, long term use of drugs
RF2102 Intra-abdominal injury
RF2201 Extensive Third-Degree Burns
RF2301 Major Organ Transplant Status

RF2302 Artificial Openings for Feeding or Elimination
RF2303 Complications of Medical & Surgical Care and Trauma
RF2304 severe morbid obesity
RF2305 morbid obesity
RF2306 obesity
RF2307 mild sleep apnea, hypoventilation
RF2308 moderate sleep apnea, hypoventilation
RF2309 obstructive sleep apnea
RF2310 Severe Protein-Calorie Malnutrition
RF2311 Mild-mod malnutrition
RF2401 Severe Head Injury
RF2402 Major Head Injury
RF2403 Vertebral Fractures without Spinal Cord Injury
RF2404 Falls, Fractures
RF2405 Amputation
RF2501 HIV/AIDS

Subtypes for PCI
STEMI
Subendocardial infarct
Unstable angina
Recent AMI
Acute CHF / pulm edema
Cardiomyopathy
Heart Failure, Cardiomegaly
Diastolic Heart Failure
Previous CABG, PCI
Heart Aneurysm and other Sequelae of AMI
Hypertensive Heart Disease
Hypertensive Heart Disease w Heart Failure
Hypertensive Heart Disease w Heart Failure & CKD
Renovascular and other secondary hypertension
Pulmonary heart disease
Sinus Node Dysfunction
Atrial Flutter / Fibrillation
Supraventricular Tachyarrhythmias
Highgrade Heart Block
Other Heart Blocks / Conduction Disorders
History of Sudden Death
Other cardiac arrhythmias
Ventricular Arrhythmias
Pacemaker, Defibrillator in place
Transplanted Heart
Severe Morbid Obesity
Morbid Obesity
Obesity
Overweight
Obstructive sleep apnea
Sleep Apnea

The prevalence of the risk factors in our reference dataset are listed in the enclosed workbook entitled NQF_PCI_all_codes_risk_adjustment 06.30.15.xls – see tab “Risk Factor Prevalence”. The output of the regression model are given in the same workbook in the tab “Risk Model’.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*

Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Please refer to the enclosed excel workbook entitled (NQF_PCI_all_codes_risk_adjustment 06.30.15.xls).

Assembling the Denominator:

Using administrative claims database, patients undergoing a PCI are identified using one of the following criteria: 1) Patients with a procedure code of PCI in any position on an in-patient of an out-patient facility claim with a qualifying diagnosis code relevant to the PCI procedure, 2) Patients having a professional service carrying a trigger code of PCI in any position. The trigger codes for PCI are provided in the tab called "Triggers I-9" or "Triggers I-10".

Patients are retained if they are 18 years of age or more, do not have a missing gender, have a complete episode time window in the database, have a maximum of 30-day enrollment gap for the entire episode time window, and have no outlier episode costs. All relevant professional, laboratory, imaging, ancillary and other claims that are incurred during the episode time window are included as part of the episode. Claims are considered relevant to PCI care if they have one of the diagnosis codes, as listed on the tab entitled Triggers I-9, Triggers 1-10, PACs I-9, PACs I-10, Typical Dx I-9, or Typical Dx I-10 in any position on the claim AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant readmissions and relevant admissions to post-acute care facilities are also included in the denominator. All relevant pharmacy claims carrying codes that match the ingredients listed in the Pharmacy tab of the enclosed workbook are also included as part of the episode.

If a patient has more than one concurrent episode, and the claim is relevant to both episodes, the claim could get multi-assigned, except in the case of procedural episodes that get carved out with respect to the index stay. So if an inpatient stay claim carried a procedure code that matched the trigger procedure code for PCI but they also had a qualifying diagnosis code for CAD (coronary artery disease), the stay claim would trigger both episodes concurrently, but get uniquely assigned to PCI and not be counted with CAD.

Once all the episodes are assembled, episodes that match the exclusion criteria, such as those with outlier costs, are flagged (those with total episode costs less than 1st percentile or greater than 99th percentile), and excluded from the final analysis.

Assembling the Numerator:

For every episode included in the denominator, services are flagged as having a PAC (potentially avoidable complication) based on the criteria listed below:

Any Index stay that has a PAC diagnosis code in any position except in the PRIMARY (principal) position is considered as having a potentially avoidable complication

Any readmission to an acute care facility 2 days or later after discharge but within 90-days post-discharge, that is relevant to PCI

Any admission to a post-acute care facility, that is relevant to PCI and has a PAC code in any position on the claim

Any other service (professional, outpatient facility, laboratory, imaging, ancillary) that is relevant to PCI and has a PAC code in any position on the claim

Relevant claims that do not have any PAC codes, and do not qualify as a PAC based on the criteria outlined above, are listed as

typical claims. All included relevant pharmacy services are flagged as typical. Patients that have even a single PAC claim are counted as part of the numerator.

Calculating the measure:

Proportion of PCI patients that have PACs is simply the ratio of patients with PACs within the PCI population and is called the PAC rate as shown in the equation below:

PAC rate = Patients with PCI that have at least one PAC claim / Total number of PCI patients

A flow chart demonstrating the series of steps and the counts of patients at each step is shown in tab entitled Decision Tree of the enclosed workbook called NQF_PCI_all_codes_risk_adjustment 06.30.15.xls

Drill Down Calculations:

Further analysis from this construct helps create actionable reports.

For example as shown in the tab labeled PAC overview, not only do we have the PAC rate for a population, we can break them down by the PAC type – type 1 being directly related to PCI and so actionable by the servicing physician, while type 2 PACs are related to patient safety failures and can be improved by process improvement. Additionally, analyzing what portion of the PACs occur during the index stay, vs. in the post-discharge period and how many are due to readmissions helps focus strategies in reducing them.

Risk Adjustment:

Once we have the observed PAC rates, we risk-adjust them for patient factors such as patient demographics, comorbidities collected historically, and for severity of illness or procedure using subtypes collected from the index stay and / or look-back period. This helps adjust for factors outside the providers control and levels the playing field for provider performance comparisons.

Unit of Analysis:
The unit of analysis is the individual episode.

Dependent Variable:
The dependent variable is a dichotomous variable indicating whether an episode had one or more claims assigned as a PAC (=1) or not (=0).

Independent Variables:
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_PCI_all_codes_risk_adjustment 06.30.15.xls

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., unstable angina). Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_PCI_all_codes_risk_adjustment 06.30.15.xls

As mentioned previously, to avoid creating perverse incentives all comorbidities and subtypes are identified prior to or at the very start of the episode. None are identified during the episode period.

Statistical Methods

We use logistic regression to model the probability of at least one PAC occurring during the episode. Only comorbidities and subtypes are included in the models as covariates if they are present in at least 10 episodes to prevent unstable coefficients. No further model building is conducted after the initial models are built. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but it does not make it a priority that all covariates in the model be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity of a episode condition, and lets each regression model determine for itself which of the factors are more significant for a specific episode. Non-significant covariates in episode models can not overly influence predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

When more than one line of business is included in the data, separate models are calculated for each sample (i.e., commercial, Medicaid etc.).

Provider Attribution and calculating PAC rates by provider:

Once episodes are constructed they are attributed to providers based on one of the various attribution rules. For PCI, episodes are attributed to the facility where the episode triggered, or, if the episode is triggered off a professional claim, it is attributed to the first facility claim that overlaps the professional trigger claim date.

Using the logistic regression technique described above, a model is developed that gives estimates for each risk factor and subtype for the patients in the population analyzed. These estimates are used to develop patient-level probabilities for the occurrence of PACs. The patient-level probability estimates are summed to construct aggregated measures (e.g., facility/provider-level). This method is similar to the methods employed by the Centers for Medicare and Medicaid Services (CMS) and endorsed by the National Quality Forum (NQF) to construct similar facility- and practice-level measures (i.e., mortality, readmissions, etc.):

1. For each provider, the number of actual observed occurrences of the outcome is summed across all attributed patients with that episode, to give the observed PAC rates for the provider.
2. Similarly adjusted probabilities from the risk adjustment models are summed across all attributed patients to give expected PACs for the provider.
3. The observed sum is then divided by the summed probabilities (O/E). This number yields whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). This calculation yields a practice-level unstandardized performance ratio.
4. To facilitate accurate comparisons of rates across units of analysis, this ratio is then standardized to the community rate using the indirect method. Specifically, the provider-level rate is multiplied by the expected community rate, calculated as the sum of adjusted probabilities for every individual in the sample across all providers in the analysis. This measure, known as the standardized rate, represents what the unit's risk-adjusted rate would be for the outcome of interest if its patient population was reflective of the of the overall community.

The formula for this calculation is as follows:

$$\text{Adj Outcome\_j} = \{(\text{SUM Observed\_ij})/(\text{SUM Prob\_ij})\} \times \{(\text{SUM Prob\_i}) / (\#\text{ of episodes})\}$$

Where individual is attributed to unit of analysis j (e.g., practice, provider, etc.)

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

Available in attached appendix at A.1

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF a PRO-PM, identify whether (and how) proxy responses are allowed.

Not applicable

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*

IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If data is missing, the case is deleted from both the numerator and the denominator

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The information is based on a two-year claims database from a large regional commercial insurer. The database has over 3.2 million covered lives and $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

The methodology can be used on any claims database with at least two years of data and a minimum of 150 patients with the index condition or hospitalization. Having pharmacy data adds to the richness of the risk-adjustment models.
The calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.
The methodology has been tested on databases of several health plans as well as on a few employer databases.

No data collection instrument was used.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Team, Facility, Health Plan, Population : National, Population : Regional, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Ambulatory Surgery Center (ASC), Hospital/Acute Care Facility, Other
If other: Across the care continuum

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2751_PCI_Testing_Reliability_Validity_HCI3.docx

NATIONAL QUALITY FORUM—Composite Measure Testing (subcriteria 2a2, 2b2-2b7, 2d)

Composite Measure Number: 2751
Measure Title: Proportion of Patients undergoing an Angioplasty (Percutaneous Coronary Intervention - PCI) Procedure that have a Potentially Avoidable Complication (during the episode time window)
Date of Submission: 06/30/15
Composite Construction:
☐Two or more individual performance measure scores combined into one score
☐ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

☐ Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.
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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. 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.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The information is based on a two-year administrative claims database from a large regional commercial insurer. The database contains medical and pharmacy claims on over 3.2 million covered lives and more than $25.9 billion in "allowed amounts" for costs.

1.3. What are the dates of the data used in testing? April 1, 2012 – December 17, 2014

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.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Clinician: Team, Pop: Nat, Reg, State	☐ other:

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

There were a total of 565 facilities in the data set. Because providers or facilities with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the reliability calculations. After this exclusion, there were 41 facilities left.

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

After exclusions (see 2b.3.1 below), there were a total of 5,898 episodes of PCI were included in the testing and analysis. Patients in these episodes were, on average, 55.6 years of age (range 26-64) and 31% were female. We did not have race information on these patients. All patients for this analysis had a trigger inpatient claim of PCI as identified in our code tables.

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.

For the reliability analysis, we restricted the data to only facilities with at least 10 attributed episodes. For risk adjustment, all episodes were used in the analysis, regardless of the facility to which they were attributed.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
None of the analyses included SDS variables.

2a2. RELIABILITY TESTING
***<u>Note</u>**: 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*)

We assessed the reliability of the measure to demonstrate that it sufficiently differentiates performance between providers using the beta-binomial method, which is applicable to measures of this type. Reliability is a measure that distinguishes the signal (the extent of performance variation between entities that is due to true differences in performance) from statistical noise. Our approach follows directly from the methods outlined in the technical report "The Reliability of Provider Profiling: A Tutorial" by J.L. Adams.

<u>Reference</u>:

Adams JL. The Reliability of Provider Profiling: A Tutorial. Rand Corporation. http://www.rand.org/pubs/technical_reports/TR653.html.

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)

The table below provides a summary of the reliability score for different minimum sample size thresholds. For complete results, refer to the workbook entitled, NQF_PCI_all_codes_risk_adjustment_06.30.15.xls, under the "Provider Attribution" tab to see facility-specific results.

Reliability Scores	Minimum # Episodes Per Facility	
	>=10	>=175
# of Facilities (%)	41 (100)	8 (20)
Median (IQR)	0.51 (0.26,0.62)	0.74 (0.70,0.83)
Range	0.11-0.85	0.69-0.85

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?*)

Reliability scores can vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or variation across patients within providers) whereas a reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (see Adams, 2009 cited above).

Scores among facilities with at least 10 episodes and scores for many were low. However, scores were consistently high among facilities with around 175 or more episodes. These results suggest that the measure achieves sufficient differentiation in performance among high volume facilities.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)

Content validity was built into the development of the definitions of potentially avoidable complications (PACs). This involved working with clinicians who are experts in their respective fields and specific to the episodes for which PACs are being measured. In particular, the clinical experts focused on whether or not a potentially avoidable complication can be deemed as such for a specific episode of care, and help defined and review all of the diagnosis and procedure codes for each PAC. The enclosed link lists clinicians who have participated in the various Clinical Working Groups (http://www.hci3.org/content/clinical-working-group-contributors). Some of the clinical experts have also participated in monthly webinars that highlight the clinical aspects of these measures (http://www.hci3.org/content/using-ecrs-providers).

In addition, we illustrate that our measure has face validity in several ways.

Beyond the up front work performed by clinical experts, the validity of the measure has also been tested in various real world settings. For example, we have presented results of claims data analyses that reveal the frequency and costs of PACs to physicians in several different healthcare systems involved in our pilot site implementations, as well as to medical directors from the employer coalitions and the health plans that provided the dataset to run the analyses. Some of these implementations include the Pennsylvania Employee Benefits Trust Fund and local provider groups and hospital, Horizon Blue Cross Blue Shield of NJ and many physicians and health systems.

In addition, we have performed dozens of analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:

- Vermont Payment Reform Commission
- Maine Health Management Coalition
- WellPoint / Anthem CT
- NY State Medicaid
- CT Medicaid
- CO All-payer Claims Database, Center for Improving Value in Health Care

These analyses and their results have influenced, and continue to influence, the development of various public reporting, payment reform and delivery system reform efforts. To-date, we have never experienced either wholesale or partial rejection of the results of analyses showing rates of PACs, which demonstrates the level of acceptability – face validity – of the measures from the payer, policymaker, employer and payer communities.

As importantly, measures of potentially avoidable complications have face-validity with consumers. In a series of focus groups, Judy Hibbard and colleagues[1] examined the impact of presenting information about price and quality of certain providers in influencing the decisions of consumers. They tested the validity of PACs as a discriminator of quality, as well as other measures of quality, and used the dollar symbol to illustrate the level of price, much like is done for restaurant reviews. When the PAC measure was used, respondents selected the providers with the lowest PAC rates with a high level of confidence in choice, and used it as a surrogate for a strong quality signal. To the contrary, when more standard measures of quality were used, consumers tended to ignore them and use price as a surrogate for quality. As such, what the researchers found is that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer “defects” – and lower price.

Finally, our measure definitions encompass several other measures that are accepted as being valid

complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs)[2], Hospital Inpatient Quality Reporting measures [3], Avoidable Readmissions [4,5], AHRQ defined patient safety indicators (PSIs) [6], NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates [7].

References:

[1] Hibbard JH, Greene J, Sofaer S, Fiminger K, and Hirsh J. An Experiment shows that a well-designed report on Costs and Quality can help consumers choose High-Value Health Care. *Health Affairs* 2012; 31(3): 560-568. doi: 10.1377/hlthaff.2011.1168

[2] CMS defined Hospital Acquired Conditions: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

[3] CMS operated Hospital Inpatient Quality Reporting Program: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

[4] Jencks SF, Williams MV, and Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. N Engl J Med 2009 (Apr); 360 (14): 1418-1428. doi: 10.1056/NEJMsa0803563.

[5] Casalino LP, Pesko MF, Ryan AM et.al. Small Primary Care Physician Practices have low rates of Preventable Hospital Admissions. Health Affairs, 2014; 33(9): 1-9. doi: 10.1377/hlthaff.2014.0434.

[6] Agency of Healthcare and Quality defined Patient Safety indicators: http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx

[7] NQF endorsed measures: Quality Positioning System: http://bit.ly/1E5ZdP7

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Not applicable.

2b2.4. 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?*)

Given the significant clinical input that went into developing the measure, the widespread use and acceptance the measure has gained among a wide variety of individuals and organizations across the health system (public and private payers, clinicians, consultants, patients, etc.) [1-13], and the parallels between this measure and other measures that are in widespread use, this demonstrates that the measure has strong face validity.

References:

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)

2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.

3. de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.

4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.

5. Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.

6. Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.

7. François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)

8. de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

9. Rastogi A, Mohr BA, Williams JO, Soobader MJ, de Brantes F. Prometheus Payment Model: Application to Hip and Knee Replacement Surgery. Clin Orthop Relat Res 2009; 467(10): 2587-2597.

10. de Brantes F and Rastogi A. Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. The Commonwealth Fund 40, publ. 2008; 1146:1-14.

11. de Brantes F, Gosfield A, Emery D, Rastogi A and G. D'Andrea, "Sustaining the Medical Home: How Prometheus Payment Can Revitalize Primary Care", Robert Wood Johnson Foundation Report, May 2009, http://www.rwjf.org/pr/product.jsp?id=42555, accessed October 2009.

12. de Brantes F, Camillus J. Evidence-informed case rates: a new health care payment model [Internet]. New York (NY): Commonwealth Fund; 2007 Apr [cited 2007 May 20]. Available from: http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=478278, Accessed Aug 1 2013.

13. Satin DJ, and Miles J. Performance Based Bundled Payments: Potential Benefits and Burdens. Available from: http://student.med.umn.edu/p4p-new/sites/default/files/MN%20Med%20Bundles%20Special%20Report%20-%20Satin.pdf, Accessed Aug 1 2013.

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

No formal exclusion testing was done since no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers.

Exclusions included exclusions of "patients" as well as "claims" not relevant to PCI care. Please refer to the enclosed excel workbook entitled (NQF_PCI_all_codes_risk_adjustment_06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
   a. If age is < 18 years
   b. If gender is missing
   c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
   d. If the episode time window extends beyond the dataset end date (this helps eliminate incomplete episodes).
   e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.
2. "Claims" are excluded from the measure based on the following criteria:
   a. If none of the diagnosis codes on the claim are on the list of relevant diagnosis codes (either typical Dx or PAC Dx) for PCI.
   b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for PCI.

2b3.2. 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*)

We started with a total PCI population of 10,177 episodes. After all the exclusions were applied, the remaining PCI population included in the analysis consisted of 5,898 episodes. As mentioned above, no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers. As such, no formal exclusion testing was done.

2b3.3. 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*)

No formal analysis was done on the impact of exclusions on performance scores.

Descriptive Explanation:

Exclusions of patients were for the following reasons. Some are for comparative purposes and some for medical reasons.

(a) Comparative Purposes:
We excluded patients that did not have complete enrollment for the entire episode time window. This was done to ensure that the database had complete information on patients to be able to create the entire episode. Including patients with only a partial episode window could distort the measure by artificially reducing the actual count of patients with PACs.

(b) Medical Reasons:
Patients with outlier costs (less than 1st percentile value or greater than 99th percentile) were considered to be different from the general pool, and excluded from both the numerator and the denominator. This is another way to ensure that episodes are complete (because incomplete episodes may have very low costs), and do not bring in random noise into the analysis due to inappropriate codes or services (high outliers).

_____

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1./S13 What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with 170 potential risk factors** and episode specific subtypes
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.1.2. If an outcome or resource use measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.2/S14. Identify the statistical risk model variables (Name the statistical method – e.g., logistic regression and list all the risk factor variables.
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_PCI_all_codes_risk_adjustment_06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., unstable angina). Subtypes are specific to each unique episode and are included in the models only if they are present at the start of the episode. Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their

corresponding codes in the enclosed workbook called NQF_PCI_all_codes_risk_adjustment_06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Candidate comorbidities and subtypes were included in the models as covariates if they were present in at least 10 episodes to prevent unstable coefficients.

2b4.2.1/S15. Detailed risk model specifications including coefficients, equations, codes with descriptors, definitions(may be attached in an Excel or cvs file)
All Risk Factors with their coefficients are detailed in the enclosed workbook called
All Risk Factors with their coefficients are detailed in the enclosed workbook called NQF_PCI_all_codes_risk_adjustment_06.30.15.xls – Please reference the tabs titled Risk Factor Prevalence and Risk Model.

2b4.3. Describe the conceptual/clinical <u>and</u> statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Risk factors are comorbidity indicators collected from historical claims before the start of an episode. These are universally applied across all episodes. This list was selected based on input from clinical experts in clinical working groups. In addition, the Clinical Working Groups identified episode specific severity markers that were called episode subtypes and they help distinguish an episode as being more severe than another.

All risk factors and subtypes must be present prior to, or at the start of the episode and are identified using diagnosis codes in the patient's historical claims.

To be included in the risk adjustment models, any risk factor or subtype must be present in at least 10 episodes. Beyond this no further model building was conducted to add or remove risk factors or subtypes from the model after it was initially run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates cannot overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
As explained above, no formal analysis was conducted to select risk factors. In fact, all potential risk factors and subtypes with a count of at least 10 episodes were retained to serve as predictors. The goal was to achieve a more complete explanatory model rather than achieve parsimony.

Please reference the tabs titled Risk Model in the NQF_PCI_all_codes_risk_adjustment_06.30.15.xls workbook to see the list of risk factors that met the selection criteria.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Not Applicable since our analysis did include SDS variables

2b4.5. 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*)

Model Development Approach

We used logistic regression to model the probability of at least one PAC occurring during the episode. The model included all covariates that were identified through the process above. No further model building was conducted after the initial model was run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates can not overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

For a more complete description of the risk adjustment approach, please see the document entitled, "PACs and Severity Adjustment Fact Sheet" that accompanies this submission.

Approach to Model Testing and Validation

To determine the validity and performance of the model, we used the split sample method to divide the patient sample randomly into: 1) the model building data set (80% of the sample) and 2) the test data set (20% of sample. The model was built using logistic regression on the first data set and then the coefficients from the development model were tested in the second dataset. Area under the curve (AUC) and the c-statistic were used to compare the predictive ability of the model in each of the data sets. Hosmer-Lemeshow Goodness-of-Fit tests and comparisons of observed to expected probabilities across risk deciles were further examined to assess the model's overall predictive accuracy.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

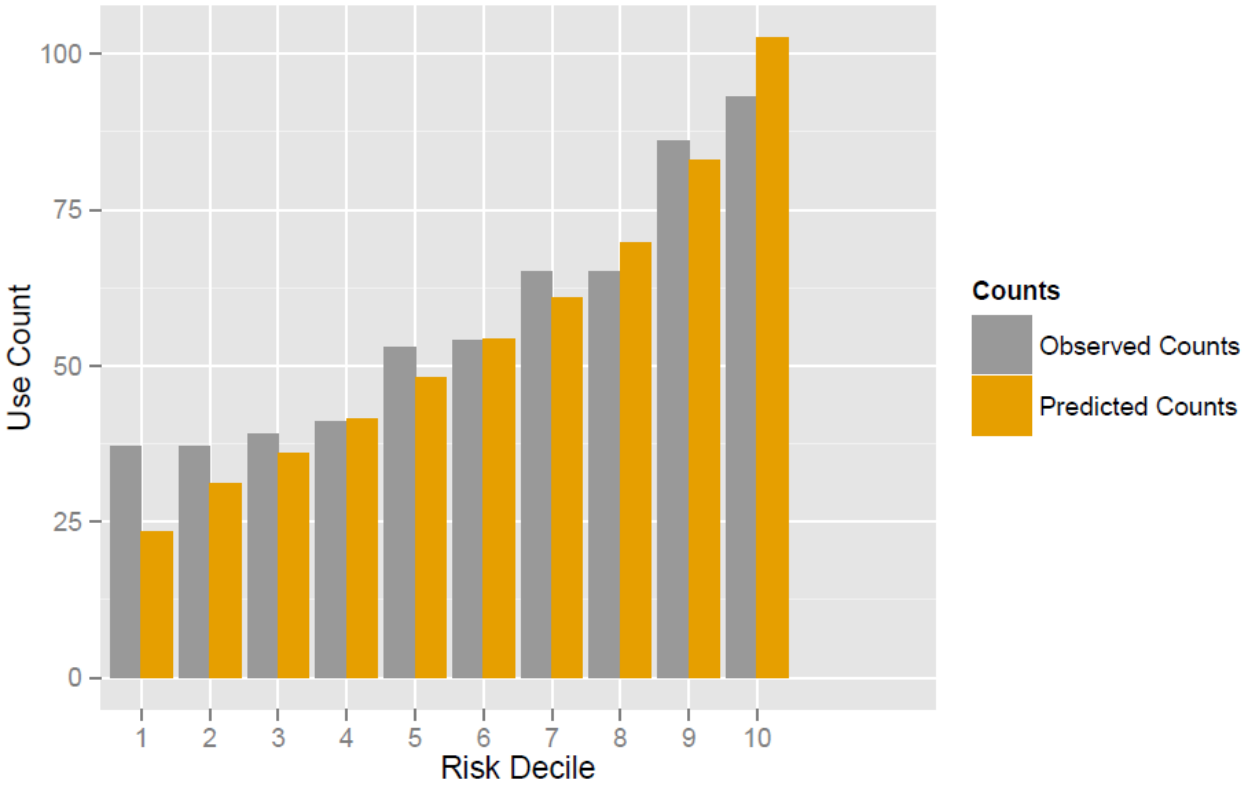
Sample	Accuracy (%)*	AUC
Test	65.9%	0.726
Validation	64.0%	0.680

*Episodes with predicted probabilities <50% were classified as having a predicted 0 (not having a PAC). Episodes with predicted probabilities >50% were classified as having a predicted 1 (having a PAC)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Sample	Chi Square	Degrees of Freedom	p-value
Test	9.7	8	0.2826
Validation	20.9	8	0.0074

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:



2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The C statistic is a measure of the extent to which a statistical model is able to discriminate between a patient with and without an outcome. The c-statistic ranges from 0.5 to 1.0. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong [1]. The purpose of the model is to adjust for patient-related factors. The remaining unexplained differences in PAC rates are due to factors that could be controlled by all providers that are managing or co-managing the patient, during the entire episode time window.

The results above indicate that the C-statistics for the risk model on the testing and validation samples (0.726 and 0.680, respectively) were around the level at which the model is considered to have good discriminatory power. Also, the accuracy values show that the model correctly predicts whether an episode had or did not have a PAC nearly 65% of the time, well above what would be expected if the predictions were made at random (i.e., 50%). The H-L test was not significant for the testing sample, meaning that the model was a good fit for the data. Finally, with the exception of the first decile, the risk decile plot shows that the model predicts PACs similarly to observed PACs across each of the other deciles.

Overall, the results demonstrate that the model has sufficient predictive power.

Reference:
[1] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Edition)*. New York, NY: John Wiley & Sons; 2000.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

NA

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To directly compare PAC rates across providers or facilities while also appropriately accounting for differences in patient severity, we calculated a risk-standardized PAC rate for each provider. This method is similar to calculations used by others for reporting outcomes measures [1]. For each provider or facility, the ratio of observed attributed episodes with PACs to the expected number of attributed episodes with PACs given the patient's risk factor and estimated from the risk-adjustment model was calculated. This number yielded whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). We then multiplied this ratio by the overall expected PAC rate across all providers or facilities to obtain the risk-standardized PAC rate for the provider or facility. This measure represents what a facility's PAC rate would be if its patient population was reflective of the overall population.

Because facilities with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. Comparison of risk-adjusted PAC rates gives a measure of the provider's relative performance. Our analysis compared risk-standardized PAC rates across facilities. We analyzed various descriptive statistics including the range in PAC rates, medians, interquartile range, etc.

References:

[1] See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

2b5.2. 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*)

Summary of Unadjusted and Adjusted Performance Scores Across Facilities:

PAC Rates	Minimum # Episodes Per Facility

	>=10	>=175
Unadjusted		
Median (IQR)	50% (44%, 56%)	47% (44%, 52%)
Range	32% - 80%	43% - 58%
Adjusted (RSPR)*		
Median (IQR)	49% (44%, 52%)	49% (46%, 52%)
Range	24% - 61%	45% - 56%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each facility.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)
The variation in risk-adjusted rates suggests there are differences in performance between facilities in risk-standardized PAC rates for patients with an episode of PCI.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
***If only one set of specifications, this section can be skipped*.**

Note: *This criterion is directed 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 SDS 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.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

2b6.2. 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*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

If patient related data is missing, the entire patient is excluded from the numerator as well as the denominator.

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence.

According to our measure definition, in constructing the measure it is possible for a provider to have only one or some types of PACs and not others. Alternatively, the provider may have all PAC types occur for their patients. The measure only considers whether any PAC occurred regardless of the type, and all PAC types are weighted equally, therefore we believe, there is no potential for the absence of specific PAC types to bias performance scores for individual providers.

For these reasons, no formal analyses were done on missing data.

2b7.2. 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*)
Not applicable

2b7.3. 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*)

Not applicable

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH
Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook entitled NQF_PCI_all_codes_risk_adjustment_06.30.15.xls provides outputs from empirical analysis. The tab labeled "PAC overview" demonstrates percentage of episodes that had at least one PAC, and provides the breakdown of PACs: 1) by the type of PAC whether directly related to index condition or due to patient safety failures; 2) the setting of the PAC, whether seen in the in-patient setting, out-patient facility or during professional visits; and 3) preventable hospitalizations.

The "PAC Drill Down Graph" provides further detail on each component of the PAC and their frequency. As can be seen by the individual counts and the graph, while each individual PAC may have such small occurrences that no meaningful comparisons in provider performances could be made; together, they add value to provide a comprehensive picture that result in meaningful numbers. The aggregation of PACs to a comprehensive, composite measure, in itself provides the parsimony that is so desirable.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

All PACs, as clinically defined by the subject matter experts were used with equal weighting. Since the emphasis of the PAC measure is to identify the occurrence of PACs in any setting, a simple and straightforward approach was adopted.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc*.; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

No formal analysis was performed.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Additionally, if a patient had one or more PACs, it is simply counted as a 1, i.e., flagged as having a PAC. The measure only considers whether any PAC occurred regardless of the type, or the site, and all PAC types are weighted equally. Therefore, no formal analysis of individual components was performed.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no*

empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each)

We chose not to weight the components of the measure.

Considerations were given to the fact that preventable hospitalizations may be given more weight, than PACs identified in a doctor's office. Similarly PACs in an in-patient setting may have more serious implications on a patient's ultimate outcome, than PACs occurring in an outpatient setting. Additionally, preventable hospitalizations as well as index hospitalizations, each with longer lengths of stay, may have serious PACs. But how do we weigh these effects? An alternative model was considered, where cost could be considered as a surrogate for the weights. Higher cost PACs could imply more serious PACs. However, differences in costs could be driven by many issues other than the PAC itself, such as unit price of the service, method of reimbursements, contracting arrangements etc.

Furthermore, in-patient facility billing does not allow for the distinction of PAC related costs from other costs within the stay. We would fail to capture PAC related costs within the stay and potentially underweight those. As a result, the decision was made to avoid weighting and keep the measure as a straightforward count.

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

Measuring all providers with the same yardstick will provide consistent results and reasonable comparisons over time. If the goal is to reduce PACs, then the PAC measure as was constructed with the help of various experts in the field would provide reasonable comparisons. A word of caution however pertains to the sample size of the provider panel before making any reasonable conclusions.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved for high reliability in one dataset would apply to another.

2d3. Empirical analysis demonstrating that the approach for handling missing data minimizes bias (*i.e., achieves scores that are an accurate reflection of quality*)**.**

Note: *Applies to the overall composite measure; the focus is on missing data rather than exclusions, which are considered in 2b3.*

Please refer to section 2b7

2d3.1. What is the overall frequency of missing data and the distribution of missing data across providers?

2d3.2. Describe the method used to compare approaches for handling missing data (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2d3.3. What were the statistical results obtained from the analysis of missing data? (e.g., *results of sensitivity analysis of effect of various rules for missing data; if no empirical analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

2d3.4. What is your interpretation of the results in terms of demonstrating that the approach used for missing data minimizes bias? (i.*e., what do the results mean in terms of supporting the selected approach for missing data; if no empirical analysis, provide rationale for the selected approach for missing data*)

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

As part of our general implementation of these measures and related analyses, we have worked through dozens of different and sometimes very large datasets. From Medicare to Medicaid to regional and national commercial carriers, as well as individual employers, the principal lesson learned is the heterogeneity of the data sets and the significant variability in fill rate of critical data elements. As a result, we have created highly specific recommendations for which data elements are required to ensure measure validity, the accuracy of those data elements, and their completeness in the dataset. When claims datasets are organized in the way we specify in the measure analysis, and contain the coding information required, the analysis of the measure and its results are highly reliable.

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 calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.

We also plan on providing a limited automated analysis, at no cost, on our website.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Payment Program Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/ Blue Cross Blue Shield of New Jersey http://www.horizonblue.com/ Pennsylvania Employee Benefits Trust Fund https://www.pebtf.org/ Quality Improvement (Internal to the specific organization) Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

4a.1. For each CURRENT use, checked above, provide:

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

Measures associated to potentially avoidable complications (PACs) are in use today with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system (public and private payers, clinicians, consultants, all-payer claims database stewards, etc.) [1-8]. They are being used in various capacities in different pilot site implementations. To name a few:

•BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
•BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers

In addition, the PAC measures are incorporated by the following organizations in their bundled payment programs:

•BCBSSC – for CABG and PCI programs
•Horizon BCBSNJ– for CHF and CABG programs
•BCBSNC
•PEBTF in PA

http://www.ajmc.com/interviews/Lili-Brillstein-on-How-Bundled-Payments-Are-Tranforming-Healthcare

In these programs they look at PACs related to the measure for process improvement activities and for practice re-engineering.

We have created reports for rates of PACs for the following organizations:
-Vermont Payment Reform
-Maine Health Management Coalition
-WellPoint / Anthem CT
-NY State Medicaid
-CT Medicaid
-CO All-payer Claims Database, Center for Improving Value in Health Care

There are several companies that are leveraging these measures to create analytics and software for customers – these include HealthQx, Aver Informatics, McKesson, and TriZetto.

Below are some references that highlight our work with Potentially Avoidable Complications (PACs).

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)
2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.
3. de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.
4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.
5. Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.
6. Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.
7. François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)
8. de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

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

4a.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.*)
Measures associated with PACs are currently in use as described in the prior section. In addition, we are working with several not-for-profit and for-profit organizations to provide them with the algorithms needed to calculate rates of potentially avoidable complications. Some of these organizations include:

Fair Health – based in NY and whose mission is to increase transparency of provider cost and quality,

CastLight – based in CA and serving large employers. We currently provide CastLight with Bridges To Excellence recognitions and will work with them to augment provider transparency by using PAC measures,

MA APCD (Massachusetts All Payers Claims Database) Council – we currently have an agreement in place with the MA APCD Council to produce PAC measures on hospitals and physicians and report back to the council with tests of reliability and validity of the measures. The purpose is to authorize the publication of these measures,

Maryland Health Care Cost Commission – we have a two year agreement to produce measures of cost and quality for public dissemination.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

We do not have any public information to share about the improvements in rates of potentially avoidable complications, as the implementation of these measures is too recent to provide valid comparisons. Further, some of the definitions of PACs have changed since the measures were initially endorsed, making comparisons even more difficult and unreliable.

Nevertheless, the variation in performance scores presented in Section 1b.2 indicates that there are differences between providers in their risk-adjusted PAC rates (higher scores equal worse performance). This suggests that real opportunities exist to identify lower performing providers and reduce the overall occurrence of PACs.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Performance results provide summary PACs rates by provider, which can be used by payers and providers in a number of ways to improve the quality of care.

From the payer perspective, payers can use this information to 1) create a high-value provider networks, 2) work with high-value providers to share best practices, 3) incentivize low-value providers to improve, 4) modify their insurance design to activate consumers to select the right care from the right providers at the right time.

From the provider perspective, providers can 1) view services and activity for their patients longitudinally across the entire care continuum, such as frequency of readmissions and ED visits and drill down on patients with high PAC rates, 2) review actionable drill down reports to identify the most frequent PACs across all patients to create care pathways and process improvement plans to impact the most frequent PACs.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended consequences were reported, but there is the potential for:
1. Under-coding of PACs in the claim stream resulting in under-reporting the actual rate and/or providers gaming the measures
2. Payers calculating the measures even with inadequate sample sizes and using the results to penalize providers

The measure is designed for transparency efforts and to spur quality improvement. Detailed PAC reports can help providers identify areas of quality improvement. Even detailed reports of small samples of patients can be helpful for quality improvement purposes, but not for public reporting. To mitigate the potential for invalid provider comparisons, we specify in this submission the minimum sample size needed to ensure the reliability of a provider's score. Ultimately, there isn't any good way to prevent provider gaming of the measure by under-coding claims, however, under the current DRG payment methodology, many providers would be penalized by under-coding PACs since these codes often result in the assignment of more complicated DRGs.

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)**

0141 : Patient Fall Rate
0202 : Falls with injury
0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0695 : Hospital 30-Day Risk-Standardized Readmission Rates following Percutaneous Coronary Intervention (PCI)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

5a. Harmonization

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 completely harmonized?

No

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

Some of the measures listed in the prior section are, fully harmonized with the submitted measure, in particular, 0705, 0708, and 0709. Other measures such as 0531, 0450, 0337, 0141, 0202 are in fact, subsets of our measure. However, there are some measures that are not harmonized, in particular the 30-day all-cause readmission measure and the Hospital wide all-cause readmission measure. While the submitted PAC measures include hospitalizations and readmissions that occur during the episode time window, the hospitalizations, by definition, have to be relevant to the index event. PACs include relevant readmissions, and are designed to enable accountability at the locus of provider control as well as some shared accountability between settings, centered around a patient, and for a specific medical episode of care. In that sense, they are consistent with the all-cause 30-day readmission rates, but represent a subset of those admissions. However, they do extend to the entire episode time window. As such, the PAC measures, as submitted, don't create added burden of reporting because the readmissions reported are simply a part of the broader 30-day all-cause readmission measures already endorsed by NQF. Because PAC measures are comprehensive, they include patient safety events that can occur during the stay, as well as adverse events, including readmissions, that can occur post-discharge. As a result, they provide facilities and physicians with an overall measure of avoidable complications for a specific medical episode. The data collection for all of the HCI3 measures is automated by a software package and is fully harmonized with

all other PAC measures. A single download automates creation of all reports related to each of the PAC measures.

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.)
PAC measures are composite measures representing "all-cause harms". They look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence. PACs look at readmissions, emergency room visits, adverse events due to errors of omission or commission. They look at complications that are due to patient safety failures, and also those directly related to the index condition. These are all a cause of significant waste and quality concerns. As such, the measure can provide clinicians with an overall and comprehensive view, in one measure, of all potentially avoidable complications for a patient and drive quality improvement efforts.

For clinicians and facilities increasingly engaged in value-based payment efforts and/or driving quality improvement for population health, the value of a PAC measure over a series of related, but more discrete measures, is that one can better determine if the sources of complications primarily stem from activities within the facility or outside the facility, and the specific nature of the complications that have a higher frequency of occurrence. While individual components of the PAC measure may have small frequencies and may be difficult to interpret with regards to provider performance or actionability, aggregating all the PACs into a comprehensive, composite measure provides the parsimony that is so desirable. For providers, it's far easier to construct a quality dashboard from a parsimonious set of measures, and that's what PAC measures offer.

Further, as a comprehensive outcome measures, PACs are also useful for public transparency of quality, as substantiated by the research from Judy Hibbard and colleagues previously cited in the "testing" section of this submission. As a comprehensive outcome measure, they are easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

Appendix

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

Attachment **Attachment:** PACs_and_Severity_Adjustment_Fact_Sheet_HCI3-635719842795809354.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.2 **Point of Contact:** Francois, de Brantes, Francois.debrantes@hci3.org, 203-270-2906-
Co.3 **Measure Developer if different from Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.4 **Point of Contact:** Amita, Rastogi, Amita.rastogi@hci3.org, 213-934-9624-

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.

From 2006 onwards, and under the auspices of various funding organizations, HCI3 has convened and managed, or helped to convene and manage, Clinical Working Groups to inform the development and refinement of the measures. For example, in 2011, 2012 and 2013, HCI3 worked collaboratively with the American Board of Medical Specialties and the American Medical Association's Physicians Consortium for Performance Improvement, under a federal contract, to convene and get input from various clinical experts on definitions of episodes of care and their sequelae, including avoidable complications.

Some of the clinical experts that have contributed to our work include:
-Dr. John Allen, American Gastroenterology Association (AGA)
-Dr. Morton Arnsdorf, Cardiologist, University of Chicago, IL
-Dr. Peter Bach, Memorial Sloan Kettering Cancer Center (MSKCC)
-Dr. Peter Basch, Primary Care, Medstar Health, DC
-Dr. Justin Beckelman, Radiation Oncology, University of Pennsylvania, PA
-Dr. Debra Bingham, Executive Director, California Maternal Quality Care Collaborative (CMQCC) at Stanford University, CA
-Dr. John Birkmeyer, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Linda Bosserman, Wilshire Oncology Medical Group, CA
-Dr. Matthew Brengman, American Society of Metabolic and Bariatric Surgery (ASBMS)
-Dr. Joel Brill, American Gastroenterology Association (AGA)
-Dr. George Cautilli, Cautilli Orthopedic Surgical Specialists PC, Yardley, PA
-Dr. Ashwini Davison, Internist, Johns Hopkins Hospital, MD
-Dr. James Denneny, III, American Academy of Otolaryngology – Head and Neck Surgery (AAO-HNS)
-Dr. Chris Gallagher, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Robert Haralson, III, American Academy of Orthopedic Surgeons (AAOS)
-Ms. Dawn Holcombe, Executive Director, Connecticut Oncology Association, CT
-Dr. Colin Howden, American Gastroenterology Association (AGA)
-Dr. John Knightly, American Association of Neurological Surgeons (AANS)
-Dr. Larry Kosinski, American Gastroenterology Association (AGA)
-Dr. Nalini Krishnan, Obstetrics & Gynecology, MN
-Dr. Kelly Kyanko, Internist, NYU School of Medicine, NY
-Dr. Tara Lagu, Internist & Infectious Disease, Baystate Medical Center, MA
-Dr. Robert Lee, Society of Thoracic Surgeons (STS)
-Dr. Alex Little, Society of Thoracic Surgeons (STS)
-Dr. Michael London, Orthopedic Surgeon, OMNI Orthopedics, OH
-Dr. Elliott Main, Obstetrics & Gynecology, California Pacific Medical Center, CA
-Dr. Constantine Mantz, 21st Century Oncology, FL
-Dr. Joseph Messer, Cardiologist, Rush University Medical Center, IL
-Dr. David Metz, American Gastroenterology Association (AGA)
-Dr. Ronald Nahass, Infectious Disease Care, NJ
-Dr. Ajay Nehra, Urologist, Rush University Medical Center, IL
-Dr. Francis Nichols, Society of Thoracic Surgeons (STS)
-Dr. Patrick O'Connor, Primary Care, HealthPartners, MN
-Dr. Sara Perkel, National Comprehensive Cancer Network, PA
-Dr. David Peura, American Gastroenterology Association (AGA)
-Dr. John Ratliff, American Association of Neurological Surgeons (AANS)
-Dr. Steven Schutzer, Connecticut Joint Replacement Institute, CT
-Dr. Leif Solberg, Primary Care, HealthPartners, MN
-Dr. Scott Sporer, Midwest Orthopedics at Rush, Chicago IL
-Dr. Bonnie Weiner, Cardiologist, Worcester Medical Center, MA
-Dr. Jonathan Weiner, Bariatric Surgery codes, Prof of Health Policy and Management, Johns Hopkins University, MD
-Dr. Janet Wright, Cardiologist, Northstate Cardiology Consultants, CA

Measure Developer/Steward Updates and Ongoing Maintenance

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

Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** Yearly
Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

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 #: 2752

De.2. **Measure Title:** Proportion of Patients undergoing Pacemaker / Defibrillator Implantation (PCMDFR) that have a Potentially Avoidable Complication (during the episode time window)

Co.1.1. **Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)

De.3. **Brief Description of Measure:** Percent of adult population aged 18 + years who had a pacemaker/defibrillator implantation (PCMDFR), are followed for at least 30-days, and have one or more potentially avoidable complications (PACs). PACs may occur during the index stay or during the 30-day post discharge period.

Please reference attached document labeled NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls, in the tabs labeled PACs I-9 and PAC I-10 for a list of code definitions of PACs relevant to PCMDFR.

We define PACs as one of two types:

(1) Type 1 PACs - PACs directly related to the index condition: Patients are considered to have a PAC, if they receive services during the episode time window for any of the complications directly related to PCMDFR, such as for wound infection, hypotension, cardiac arrest etc.

(2) Type 2 PACs - PACs suggesting Patient Safety Failures: Patients are also considered to have a PAC, if they receive services during the episode time window for any of the complications related to patient safety failures such as for sepsis, infections, phlebitis, deep vein thrombosis, pressure sores etc.

All readmissions in a patient with PCMDFR are considered potentially avoidable and flagged as PACs.

PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook labeled NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls serves as an example. The tab labeled PAC overview gives the percent of PCMDFR episodes that have a PAC and the tab labeled "PAC drill down" gives the types of PACs and their frequencies in PCMDFR episodes within this dataset.

The information is based on a two-year claims database from a large regional commercial insurer. The database had over 3.2 million covered lives and over $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

1b.1. **Developer Rationale:** Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability (McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 40% of its plan members with hypertension incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement

improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow- up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.

4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

6) BCBSNC: Blue Cross Blue Shield of North Carolina:
https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5.
http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. "The Prometheus Model: Bringing Healthcare into the Next Decade." Annals of Health Law Advance Directive 19 (2010): 274-284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/

11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. "An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care." Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.

14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

15) Miller MR, Elixhauser A, Zhan C, and Meyer G. "Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns." Heath Services Research 36.6.2 (2001): 110-132.

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

S.4. **Numerator Statement:** Number of patients who underwent a pacemaker/defibrillator implantation (PCMDFR), are followed for at least 30-days, and have one or more potentially avoidable complications (PACs) during the episode time window.
S.7. **Denominator Statement:** Adult patients aged 18 years and above who underwent a Pacemaker/defibrillator implantation - PCMDFR) procedure and are followed for at least 30-days.
S.10. **Denominator Exclusions:** Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for any time period during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the PCMDFR measure if they are considered not relevant to PCMDFR care.

De.1. **Measure Type:** Composite, Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Clinician : Individual, Clinician : Group/Practice, Clinician : Team, Facility, Integrated Delivery System
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

IF this measure is included in a composite, NQF Composite#/title: n/a – 2751 is i an "any or none" composite measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)
Composite Measure Construction: n/a. The individual complications are considered measurable components

Is this a MAINTENANCE measure submission? ☐ **Yes** ☒ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: n/a **Most Recent Endorsement Date:** n/a

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating clinical evidence asks if the there is a relationship between the measured health outcome and at least one clinical action is identified and if it is supported by the stated rationale. For a composite measure, the developer must discuss the reasoning for the

composite quality constructs, the rationale for constructing, & aggregation and weighting of measure components.

- This new risk-adjusted (by age, gender and clinical co-morbidities) outcomes composite measure assesses the proportion (rate) of adult patients undergoing Pacemaker / Defibrillator Implantation (PCMDFR) with *at least one* Potentially Avoidable Complications (PAC) for the measure time window.
- Based on NQF's criteria, this measure is considered an "any or none" composite measure that assesses if 1 or more PACs or "care defects" have occurred for the index episode. For this composite measure, the individual complications considered the measurable components. PACs are classified in two types: 1) related to PCMDFR, and 2) related to Patient Safety Failures. The2 PAC types are combined into a single "any or none" (bimodal "yes" or "no") PAC rate. PACs are considered unwarranted health outcomes that combine concepts from AHRQ PSIs, PQIs and the CMS HACs and episode-specific PACs into index episode all-cause patient harms rate.
- The developer links errors of commission/omission (poor safety practices) to unnecessary ER visits, hospitalizations, readmissions, and mortalities to increased PACs. The developer further states that PACs for PCMDFR patients should occur rarely in well-managed patients, and defines potential avoidable PCMDFR complications, that include (but are not limited to) line sepsis, infections, pocket hematomas, pneumothorax, perforations, and death.
- The evidence for Patient Safety Failure PACs is described to be within the influence of the measured entity, though the rationale for selecting some of the identified PACs is not clear (e.g., post procedural fever, oral bisphosphonates, hallucinations). The developer provides an extensive list of comorbidities as risk factor for increased PAC potential, though the severity is not captured in consistently within the claims data.
- In addition to linking processes of care to outcomes, the developer provides an extensive PAC literature review in sections 1a.2. and 1a.2.1. for PCMDFR, Patient Safety Failures & processes of care, as well as background information on the process for PAC development.
- The developer discusses the rationale for constructing, aggregation and equal weighting for the measure.

Questions for the Committee:

- *Does sufficient evidence exist connecting Patient Safety Failures to the PCMDFR index episode?*
- *Does sufficient evidence exist between the measured health outcome and at least one clinical action identified and supported by the stated rationale?*

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

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

- PCMDFR PCA performance gap data are calculated from PROMETHEUS administrative claims data from April 1, 2012 through December 17, 2014, for 1,806 of 3,968 (45.5%) PCMDFR episodes (in 3,258,706 unique beneficiaries) and 22 of 380 (5.8%) facilities after excluding fewer than 10 attributable episodes due to unstable small samples.

Unadjusted PAC Rates:		**Risk-Standardized PAC Rates (RSPR):**	
Median (IQR):	46.8% (39.5%, 55.2%)	Median (IQR):	46.2% (36.8%, 54.0%)
Range:	20% - 64.3%	Range:	20.8% - 62.5%

- Descriptive data on the patient, facility and payer are not provided. The developer provides "Overview" and "Drill Down" PAC rates in the spreadsheet demonstrating gap, though PAC rates for individual complications are not provided.
- The developer does not provide data on disparities.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - There is clear evidence that measures that can minimize complications of pacemaker and ICD implants are important. However, I have major concerns about the measure as it is proposed.

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- It is not clear what would be considered a PAC or not. Where do they draw the line? What about HF? A stroke? An MI not related to the procedure? What about trauma related to MVA? Do ICD shocks count as well? What if the programming of the device is in line with the evidence why would we penalize clinicians for that?

1b. ***Committee's Comments on Performance Gap:***

- PCMDFR PCA performance gap data are calculated from PROMETHEUS administrative claims data from April 1, 2012 through December 17, 2014, for 1,806 of 3,968 (45.5%) PCMDFR episodes (in 3,258,706 unique beneficiaries) and 22 of 380 (5.8%) facilities after excluding fewer than 10 attributable episodes due to unstable small samples.
- Unadjusted PAC Rates:
  - Median (IQR): 46.8% (39.5%, 55.2%)
  - Range: 20% - 64.3%
  - Risk-Standardized PAC Rates (RSPR):
  - Median (IQR): 46.2% (36.8%, 54.0%)
  - Range: 20.8% - 62.5%
- Descriptive data on the patient, facility and payer are not provided. The developer provides "Overview" and "Drill Down" PAC rates in the spreadsheet demonstrating gap, though PAC rates for individual complications are not provided.
- The developer does not provide data on disparities.

1c. ***Committee's Comments on Composite Performance Measure:***

- Any or none measure.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The measure assesses the rate of patients with 1 or more PAC(s) during index episodes. This new risk adjusted outcomes measure is specified for use at the individual clinician, group/practice, team, facility & integrated delivery system levels of analyses.
- The measure exclusively uses electronically available administrative claims data to calculation the measure score, and for this measure, better care equals lower scores.
- The developer describes non-patient-related PACs as controllable by facility processes without further analysis as other influencers that may contribute to PAC rates beyond the patient and facility (e.g., payer, access, suppliers, etc.).
- Patient- and claims-based exclusions are provided to promote the availability and consistency of claims data capture, including payer enrollment requirements, cost outliers of < 1% or > 99%, and claims not relevant to PCMDFR.
- Developers provide administrative claims codes for PCMDFR & PAC (PCMDFR - & Patient Safety Failure-related) triggers, and describe a 7 day look back and 30 days after the PCMDFR triggered claims data. The developer should

provide context for the use of claims data prior to the PCMDFR in relation to triggering index episodes and/or PACs.

- A calculation algorithm is provided, as well as ICD-9 & ICD-10 codes, though ICD-10 descriptions & an ICD-9 to ICD-10 crosswalk methodology are not provided.
- A conceptual risk model and statistical method using logistic regression model for determining the probability of a patient incurring a PAC are provided. After adjusting for patient-related factors, the developers state the remaining PAC variance is due to factors potentially controlled by the facility during and after hospitalization. "Predicted" coefficients from the risk adjustment models are summed to give predicted probabilities of PAC occurrence.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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 facilities.

- The developer tested reliability at the performance measure score, and used a beta-binomial model and a signal-to-noise analysis, which is appropriate for this type of measure, to differentiate the true difference between measured entities (the signal) to random measurement error (the noise). A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in between *facility* performance. A value of 0.7 is often regarded as a minimum acceptable reliability value, and the developer also states values above 0.9 are considered sufficient to see differences between pairs of physicians.
- The measure is specified for patients ≥ 18 years that underwent PCMDFR, though the testing sample includes patients 18 through 64 years.
- The measure is specified for use with individual clinician, group/practice, team, facility & integrated delivery system levels of analyses, though testing is provided for facilities. NQF's measure evaluation criterion requires testing for all measure specification levels.
- Facilities with < 10 PCMDFR episodes were excluded from reliability testing, though the measure is specified for patient without episode restrictions. A sample of 380 facilities was initially included in the data set, though facilities with less than 10 PCMDFR episodes were excluded, allowing for 22 (5.8%) remaining facilities. There were 1,806 episodes of PCMDFR with a mean age of 54.1 (18-64 years) with 31% being female.
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- The developer states, "Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset will apply to another." The developer also states that very high sample sizes are to achieve any meaningful and reliable comparisons.
- In section 2a2, the developer states, "The beta-binomial failed to produce statistically significant parameters. We were therefore unable to calculate facility reliability scores. We were unable to report reliability scores, suggesting that statistically the measure may not adequately differentiate between facilities in the current database tested."
- Due to the original reliability testing results, an ad hoc commercial data set demonstrating "providers with sample sizes as low as 22 patients, had a reliability >0.7", though specific results are not provided (details are provided in the Ad Hoc Pacemaker Defibrillator Reliability Analysis spreadsheet in SharePoint).
- A patient may have more than one condition-specific concurrent episode with claims applied to both episodes. If an inpatient claim corresponds to a procedure index episode and to a condition index episode, the claim would be assigned to the procedure index episode, rather than the condition index episode (e.g., for a claim that corresponds to both index episodes of CAD & CABG, the claim would be assigned to CABG).
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- The developer provides an additional supplementary fact sheet related to PAC development & testing (available for

review on SharePoint).

Questions for the Committee:

- *Reliability testing was attempted only for those facilities with at least 10 episodes. Can differences in performance be identified for facilities with fewer than 10 or 22 episodes? For patients ≥ 65 years?*
- *Should the measure be specified to include only those facilities with at least 10 episodes?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- Because this is an outcome measure, the rationale that is presented for subcriterion 1a does not necessarily have to address all of the variables used to calculate the measure.
- The measure uses a statistical risk model with 170 risk factors and episode-specific subtypes/severity markers including age, gender and clinical comorbidities, on at least 10 claims to determine "stable" covariates and assess comorbidity or procedure impact on the PAC. All covariates must be present prior to an episode trigger. No formal covariate analysis was conducted to select risk factors beyond the minimum of 10 claims threshold. The developer describes the heterogeneity of the provided data sets as crucial to ensure measure validity, and the accuracy and completeness of the data sets.
- The developers did not provide disparities data, an exploration of a conceptual relation to SDS, or SDS factors in the risk model.

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Are these variables available and generally accessible for the measured patient population?*

2b2. Validity testing

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.

- The developer conducted systematic assessment of face validity for the performance measure score for validity testing in numerous ways, including the use of monthly multi-specialty clinical working groups, and other tests of face validity, along with focus groups, face validity comparisons of the measure to other national accountability measures, as well as additional literature for the measure & PAC development process.
- No empiric results are provided for the face validity tests described above.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer describes patient- (demographic, enrollment or low/high claims cost) and claims-based (due to missing or non-relevant data) exclusions for the measure. They further state nearly half of the original population of PCMDFR

patients was removed from the denominator with applied exclusions.

- A significant number of episodes were eliminated from the measure due to exclusion criteria, permitting 1,806 of 3,968 (45.5%) PCMDFR episodes (in 3,258,706 unique beneficiaries) and 22 of 380 (5.8%) facilities for analysis.

Questions for the Committee:

- *Are high cost outliers (> 99%) exclusions an opportunity to identify PACs?*
- *Does the high number of exclusions restrict the measure use?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across facilities to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- The risk model (detailed in the accompanied spreadsheet) includes 170 factors and subtypes including age, gender, 12-month enrollment markers, co-morbidities, and episode severity markers.
- No SDS factors beyond age and gender were included in the risk-adjustment approach. The developers note that race was not available for analysis, and no description of the of the conceptual relationships between patient sociodemographic factors, patient clinical factors, quality of care, and the outcomes (PAC rates) were provided, nor do they discuss the availability of SDS variables.
- Logistic regression was used to model the probability of at least one PAC during an episode. The reasoning for no additional modeling performed is described.
- The performance of the model was determined with a split sample method by estimating the model coefficients using a development dataset (80% of the sample) and applying those coefficients to a validation dataset (20% of the sample). C-statistics for the development and validation samples with c-statistic results of 0.740, for the test sample only. C-statistics measures the extent of a statistical model to discriminate between a patient with and without PAC, with an ability to predict if a PAC is or is not present about 60% of the time. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong.
- Both Hosmer-Lemeshow Goodness-of-Fit statistics and risk-decide plots were provided to indicate model fit. Results from the Hosmer-Lemeshow test suggest that the fit is not good; however, this test is sensitive to the number of groupings and sample sizes. Results from the risk decile plot indicate that the predicted PAC rates are similar to the observed PAC rates across all deciles of risk. The developer states the model demonstrates strong predictive power.

Questions for the Committee:

- *Is the Committee aware of conceptual relationship(s) between additional patient-level SDS factors, patient clinical factors, quality of care, and PACs (other than gender and age)? If so, what data might be available to allow an empirical analysis of these relationships?*
- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*

2b5. Meaningful difference:

- The developer presents PAC rates across facilities and also facilities adjusting for differences in patient severity in a ratio of observed to expected attributable episodes to PACS accounting for patient severity, and calculates estimates from the risk model, for risk-standardized PAC rates for the facility.
- The table below provides PAC rates per provider. The developer should clarify PACs in the table represent individual providers or facilities.

Summary of Unadjusted and Adjusted Performance Scores Across Facilities:

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25
# Providers	22	16
Unadjusted		
Median (IQR)	47% (40%, 55%)	46% (35%, 55%)
Range	20% - 64%	26% - 64%
Adjusted (RSPR)*		
Median (IQR)	46% (37%, 54%)	46% (35%, 55%)
Range	21% - 63%	25% - 63%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_ PCMDFR _all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each provider.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- As there is only one data source used for measure calculation (administrative claims), comparability of data sources or methods is not applicable.

2b7. Missing Data

- No formal analysis of missing data is provided. As the measure assesses the rate of patients with PACs, rather than the rate of PACs per index episode, the total number of PACs is not included in the PAC rate.
- Patient with missing gender were excluded from the denominator, and no other missing data was identified.
- The developers state the under-coding of claims is unavoidable in the current DRG payment structure which could lead to under capture or missing PACs.

2d. Empirical Analysis to Support Composite Construction

- As an "any or none" composite, the individual complications are considered measurable components of the composite. Frequency and distribution statistics are provided in the PAC Overview and PAC Drill Down tabs in the measure spreadsheet, which detail PAC types and subtypes. The identification of individual PACs are not provided (e.g., sepsis, unattended falls, DVT).
- PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.
- The developer states that no formal analysis was performed on missing data. For details, see 2b7 above.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- Under exclusions: What about CRT? What about device replacements or procedures only involving leads etc.. Otherwise, ok.

2a2.: ***Committee's Comments on Reliability-Testing:***

- The sponsor was not successful in proving the following: The measure is specified for use with individual clinician, group/practice, team, facility & integrated delivery system levels of analyses, though testing is provided for facilities. NQF's measure evaluation criterion requires testing for all measure specification levels.
- The following is concerning: "The beta-binomial failed to produce statistically significant parameters. We were therefore unable to calculate facility reliability scores. We were unable to report reliability scores, suggesting that statistically the measure may not adequately differentiate between facilities in the current database tested."
- I am not sure I understand this: Due to the original reliability testing results, an ad hoc commercial data set

demonstrating "providers with sample sizes as low as 22 patients, had a reliability >0.7", though specific results are not provided.

2b1.: ***Committee's Comments on Validity-Specifications***:

- Not Applicable

2b2.: ***Committee's Comments on Validity-Testing:***

- I am worried that one would have to adjust for 170 factors. Why that many? How were these chosen? May did not appear to be relevant. Also, with the need to adjsut for 170 factors, is the measure even practical or feasible to use?
- No empiric results are provided for the face validity tests described above.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- The developer describes patient- (demographic, enrollment or low/high claims cost) and claims-based (due to missing or non-relevant data) exclusions for the measure. They further state nearly half of the original population of PCMDFR patients was removed from the denominator with applied exclusions.
- A significant number of episodes were eliminated from the measure due to exclusion criteria, permitting 1,806 of 3,968 (45.5%) PCMDFR episodes (in 3,258,706 unique beneficiaries) and 22 of 380 (5.8%) facilities for analysis. These are concerning.

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than the person obtaining the original information (e.g., DRG, ICD-9 codes on claims).
- The developer provides an excel spreadsheet attachment including diagnoses, visits, hospitalizations, post-acute facility stays, procedures, laboratory tests and procedures/surgeries, for PCMDFR & PAC triggers, and describe the time window for measuring PAC triggers as a 7 day look back and 30 days after undergoing a PCMDFR, as well as a decision tree for measure calculation and implementation.
- The measure specifications, metadata and calculation algorithms are available free of charge on the developer's website. Limited analytics are planned at no cost to the end user.
- This is not an eMeasure.

Questions for the Committee:

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- I am not sure

Criterion 4: Usability and Use

4. Usability and 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.

- This is a newly developed claims measure is current used in accountability programs for payers, states, and planned for public reporting, professional certification or recognition programs, and external quality improvement for benchmarking purposes.
- The developer states that PAC measures provide a foundation for the relationship between healthcare quality and cost and assist in the exploration of practice reengineering and alternative payment models, act as indicators of potential harm, and is spurring the development of private-based analytics software for further outcomes exploration. No public improvement rates are available due to recent implementation and variation in PAC definitions have also modified, though the provided PROMETHEUS data suggest wide variation in performance and improvement opportunities.
- The developer found no noted unintended consequences, though they state the measure is intended for transparency and QI activities only. They also state the under-coding of claims is unavoidable in the current DRG payment structure could be an unintended consequences of the measure, and payers calculating the measures even with inadequate sample sizes and using the results to penalize providers could lead to invalid provider comparisons.
- If the measure was theoretically to be used for accountability purposes to "ding" the measured entity as defined in the level of analysis, further exploration of PAC antecedents and the measured entity is warranted, especially with lower volume PCMDFR facilities. In the original testing sample of 380 facilities, when facilities with fewer than 10 PCMDFR episodes were eliminated from analysis due to less reliability estimates with small numbers, 22 (5.8%) remained for analysis.

***Questions for the Committee**:*

- *Is the measure publicly reported?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Should PAC measures also include the clinician: group in the analysis or include population-level only entities?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- I am not convinced.

Criterion 5: Related and Competing Measures

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

0141 : Patient Fall Rate
0202 : Falls with injury
0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2752

Measure Title: Proportion of Patients undergoing Pacemaker / Defibrillator Implantation (PCMDFR) that have a Potentially Avoidable Complication (during the episode time window)

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: 6/30/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

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

. Evidence to Support the Measure Focus

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

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☒ Health outcome: Potentially Avoidable Complications

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

With the expanded indications for Pacemaker / Defibrillator (PCMDFR) use, improved technology and increasing number of patients with cardiac diseases, the use of PCMDFR's has increased exponentially (van Rees 2011) (Greenspon 2012). Dedicated quality improvement efforts to enhance patient safety, and reduce procedural complications are required to be in place, and could pay for themselves with better patient outcomes and reduced costs (Reynolds 2006). PCMDFR complications could be reduced by a combination of optimal medical treatment as well as adoption of adequate implantation techniques for these devices (van Rees 2011). Optimal placement requires a clinical center, hosting a team of qualified and experienced cardiologists, nurses and technicians. The teams' experience, and the volume of pacemakers implanted in the

center, plays a role in reducing post-implantation complications. Studies have shown low volume centers (<750 procedures per annum) to have 50-100% higher risk of any complication compared to high volume centers (Kirkfeldt 2013).

Potentially avoidable complications (PACs) are the health outcomes that this measure addresses. PACs are both directly and indirectly related to healthcare services provided (or not provided) for a condition or procedure (de Brantes 2010). PACs may occur due to errors in omission or commission. Errors of omission in a patient undergoing a pacemaker / defibrillator implantation could be due to failure of hospitals and / or physicians to establish or implement patient safety protocols when inserting pacemaker leads or the defibrillator assembly leading to line sepsis, infections etc. (Pronovost 2010). In addition, errors of commission could be due to improper placement of leads resulting in pneumothorax, perforation during PCMDFR lead placement or lead dislodgement as a late complication (van Rees 2011). Pocket hematomas, if not diagnosed early could lead to a 15-fold increase in wound infections (Klug 2007). Lack of operator experience, lack of care coordination, poor discharge planning and poor arrangements of patient follow-up could lead to unnecessary ER visits, readmissions and gaps in care leading to increased morbidity and the need for lead replacement. Cumulative incidence of inappropriate shocks can increase to 18% and results in an increased all cause mortality (van Rees 2011). Readmissions are common after inappropriate shocks for reasons varying from battery failure, addition of antiarrhythmic medications, treatment of associated MI, to treatment of discomfort and psychological support for anxiety (Beyerbach 2014). Minimizing inappropriate defibrillator shocks is therefore of paramount importance and requires a multifactorial approach including, appropriate patient selection, medical care, counseling, choosing the right device and appropriate programming to reduce shocks (Koneru 2011). Up to 35% of patients could develop anxiety disorders following implantation and need appropriate psychosocial support (Kamphuis 2003). All these adverse events are aggregated together in the PAC measure to study the overall rate of PACs in the PCMDFR population.

Adult patient undergoing a pacemaker or defibrillator implantation admitted to a hospital or an outpatient facility

↓

Hospital/physician fails to carry out safe practices (errors in commission/omission)

↓

Patient suffers complication stemming from hospital/physician potentially avoidable error

↓

Patient remains in hospital for treatment of PAC

OR

Patient readmitted to hospital with 1+ Potentially avoidable complication

Well-managed patients receieving a pacemaker or defibrillator should rarely incur a potentially avoidable complication such as an emergency room visit post-discharge, and readmissions related to PCMDFR should occur only in the rarest of circumstances.

The enclosed workbook entitled NQF_PCMDFR_all_codes_risk_adjustment 06.30.15.xls lists the types of PACs and their frequency as calculated in a large regional database (see tab PAC overview). Over 47% of PCMDFR episodes had a PAC. Of these, over 37% were incurred for direct complications of PCMDFR, such as malfunction / complications of the device, or lung complications from line insertion (see tab PAC Drill Down Graph). Although the preventable readmissions in the PCMDFR population were low, at only 3.8%, approximately 20% of patients with PCMDFR had PACs related to patient centered care failures such as respiratory insufficiency and poor control of diabetes, many of them being managed in an outpatient setting in physician offices. As a result 42% of episodes had a PAC indicator on the professional claims.

While PACs may not be eliminated completely, identifying the magnitude of PACs and knowledge of the cause for the most frequent or the most expensive PACs could place an emphasis in reducing them and as a consequence improving patient outcomes. The ability to clearly identify the type and frequency of each PAC creates a highly actionable measure for all providers that are managing or co-managing the patient, as well as for the health plan with whom the patient is a member.

References:

1) Van Rees, J. B. "Inappropriate Implantable Cardioverter-defibrillator Shocks: Incidence, Predictors, and Impact on Mortality." *J Am Coll Cardiol* 57.5 (2011): 556-62. *NCBI*. Web.

2) Greenspon, Arnold J., Jasmine Patel, Edmund Lau, Daniel Frisch, Reginald Ho, Behzad Pavri, Jorge Ochoa, and Steven Kurtz. "Trends In Permanent Pacemaker Implantation In The United States 1993-2009: Increasing Complexity Of Patients And Procedures." *Journal of the American College of Cardiology* 59.13 (2012). Web.

3) Reynolds, Matthew R., David J. Cohen, Aaron D. Kugelmass, Phillip P. Brown, Edmund R. Becker, Steven D. Culler, and April W. Simon. "The Frequency and Incremental Cost of Major Complications Among Medicare Beneficiaries Receiving Implantable Cardioverter-Defibrillators." *Journal of the American College of Cardiology* 47.12 (2006): 2493-497. Web.

4) van Rees, Johannes B., Mihály K. De Bie, Joep Thijssen, C. Jan Willem Borleffs, Martin J. Schalij, and Lieselot Van Erven. "Implantation-Related Complications of Implantable Cardioverter-Defibrillators and Cardiac Resynchronization Therapy Devices." *Journal of the American College of Cardiology* 58.10 (2011): 995-1000. *Pubmed.gov*. Web.

5) Kirkfeldt, R. E., J. B. Johansen, E. A. Nohr, O. D. Jorgensen, and J. C. Nielsen. "Complications after Cardiac Implantable Electronic Device Implantations: An Analysis of a Complete, Nationwide Cohort in Denmark." *European Heart Journal* 35.18 (2013): 1186-194. Web.

6) de Brantes, F., A. Rastogi, and M. Painter. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach". *Health Services Research* 45.6.2 (2010): 1854-1871.

7) Pronovost, P.J., G. A. Goeschel, E. Colantuoni et al., "Sustaining Reductions in Catheter Related Bloodstream Infections in Michigan Intensive Care Units: Observational Study," *BMJ* (Feb. 4, 2010): 340:c309. Web.

8) Klug D, M. Balde, D. Pavin, et al. "Risk factors related to infections of implanted pacemakers and cardioverter-defibrillators: results of a large prospective study." *Circulation* 116 (2007): 1349 –55.

9) Beyerbach, Daniel M., MD, PhD, and Jeffrey N. Rottman, MD. "Pacemakers and Implantable Cardioverter-Defibrillators." Pacemakers and Implantable Cardioverter-Defibrillators. Medscape, 2 May 2014. Web. <http://emedicine.medscape.com/article/162245-overview>.

10) Koneru, J. N., C. D. Swerdlow, M. A. Wood, and K. A. Ellenbogen. "Minimizing Inappropriate or "Unnecessary" Implantable Cardioverter-Defibrillator Shocks: Appropriate Programming." *Circulation: Arrhythmia and Electrophysiology 4*.5 (2011): 778-90. Web.

11) Kamphuis HC, et al. "Implantable cardioverter defibrillator recipients: quality of life in recipients with and without ICD shock delivery: a prospective study." *Europace* 5.4 (2003 Oct): 381-9.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

Rationale: There exists a continuing need for reduction of the clinical and economic burden of PAC's due to PCMDFR's. PCMDFR use is associated with complications at all stages of device use, starting from in-hospital stay to post discharge. Associated comorbidities, type of device used as well as physician volumes further influence the complication rates. Various studies have shown that the commonest complications like lead displacement, pneumothorax, infection and perforation are all PAC's and are influenced by physician experience and training. Complex lead placements are usually performed by more experienced consultants

and consequently had lower displacement rates (Bond 2012). Implantations performed in an emergency or out-of-hours settings are also associated with higher complication risks (Kirkfeldt 2013).

Shocks from ICD's though life saving in patients paradoxically can also cause much of the morbidity associated with their use and reduce quality of life. Cumulative incidence of inappropriate shocks can increase to 18% and results in an increased all cause mortality (van Rees 2011). Readmissions are common after inappropriate shocks for reasons varying from battery failure, addition of antiarrhythmic medications, treatment of associated MI, to treatment of discomfort and psychological support for anxiety (Beyerbach 2014). Minimizing inappropriate ICD shocks is therefore of paramount importance and requires a multifactorial approach including, appropriate patient selection, medical care, counseling, choosing the right device and appropriate programming to reduce shocks (Koneru 2011). Up to 35% of patients can develop anxiety disorders following implantation and need appropriate psychosocial support (Kamphuis 2003). Appropriate programming could minimize unnecessary shocks in patients receiving an implantable pacemaker / defibrillator (Koneru 2011), in turn reducing unnecessary anxiety disorders in these patients (Kamphuis 2003).

Mariana Parahuleva, in her extensive review of PCMDFR complications published as a book chapter states that, identification of factors contributing to complications may permit identification of high-risk individuals that warrant incremental monitoring and therapy to attenuate risk (Parahuleva 2011). Centers should strive for significant reductions in frequency of complications related to pacemaker/defibrillators through adopting and incorporating technological developments, improved operator competence and patient education.

PAC measures in the setting of pacemaker / defibrillator (PCMDFR) implantation look at all-cause harms, such as the ones highlighted above, arising from poor management of a patient receiving a PCMDFR.

References:

1) Bond, Richard, Daniel Augustine, and Mark Dayer. "Pacemaker Complications in a District General Hospital." British Journal of Cardiology Br J Cardiol 19.2 (2012): 90-94. Web.

2) Kirkfeldt, R. E., J. B. Johansen, E. A. Nohr, O. D. Jorgensen, and J. C. Nielsen. "Complications after Cardiac Implantable Electronic Device Implantations: An Analysis of a Complete, Nationwide Cohort in Denmark." *European Heart Journal* 35.18 (2013): 1186-194. Web.

3) Van Rees, J. B. "Inappropriate Implantable Cardioverter-defibrillator Shocks: Incidence, Predictors, and Impact on Mortality." *J Am Coll Cardiol* 57.5 (2011): 556-62. *NCBI*. Web.

4) Beyerbach, Daniel M., MD, PhD, and Jeffrey N. Rottman, MD. "Pacemakers and Implantable Cardioverter-Defibrillators." Pacemakers and Implantable Cardioverter-Defibrillators. Medscape, 2 May 2014. Web. <http://emedicine.medscape.com/article/162245-overview>.

5) Koneru, J. N., C. D. Swerdlow, M. A. Wood, and K. A. Ellenbogen. "Minimizing Inappropriate or "Unnecessary" Implantable Cardioverter-Defibrillator Shocks: Appropriate Programming." Circulation: Arrhythmia and Electrophysiology 4.5 (2011): 778-90. Web.

6) Kamphuis HC, et al. "Implantable cardioverter defibrillator recipients: quality of life in recipients with and without ICD shock delivery: a prospective study." *Europace* 5.4 (2003 Oct): 381-9.

7) Parahuleva, Mariana. "8 Cardiovascular Implantable Cardioverter Defibrillator-Related Complications: From Implant to Removal or Replacement: A Review." Cardiac Defibrillation - Mechanisms, Challenges and Implications. Giessen, Germany: INTECH Open Access, 2011. Web.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form 2752_PCMDFR_Evidence_Attachment_HCI3.docx
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., the benefits or improvements in quality envisioned by use of this measure)* Measures associated to potentially avoidable complication (PAC) have been used as comprehensive outcomes measures since 2007 for several conditions and procedures (de Brantes 2010) (Joynt 2013) (James 2013). In 2011, following the NQF endorsement of these measures for certain acute medical conditions (AMI, Pneumonia and Stroke), and for chronic conditions, they were adopted for various purposes, including the creation of related measures (NQF – Measure #1550). Some commercial payers have used them as a means for tracking outcomes (Yong 2010) and for tiering providers for pay for performance programs (BCBSNC). In addition, some provider organizations have used them in quality improvement efforts by homing in on the detailed specifications of the measures to reveal opportunities for care improvement (CALPERS – link below). Identification of PACs has spurred provider innovation (Bundled Payment Summit 2015) for practice re-engineering, to create proactive care pathways, and to focus on areas of high variability

(McVary 2010). Some employers are also using measures of avoidable complications as public measures of quality (Colorado Business Group on Health) given the research that demonstrated the potential efficacy of these measures to differentiate provider quality and cost (Hibbard 2012). In fact in a series of focus groups led by Judy Hibbard and colleagues, the researchers found that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Accountability for and measurement of PACs occurs at the practice, medical group, provider system or purchaser/payer level. PAC rates are calculated as absolute values. For example, a health plan would report that 40% of its plan members with hypertension incurred PACs in the study time window. The objective of the measure is to encourage the unit being measured to progressively reduce that amount over time. In addition, comparisons of PAC rates across plans or provider systems should be encouraged and publicly reported. An organization that uses the measure should be able to identify the leading causes of PACs and implement improvements to existing processes that will decrease PACs. There are several tools available for provider systems and health plans to impact PAC rates. These include care coordination across care settings; post-discharge planning and patient follow- up, active care management, sharing medical record data between care settings and providers, total quality management within hospitals and active reduction of patient safety failures. Reducing PACs has the potential to significantly improve the overall level of quality.

Creating a single measure of accountability for physicians and hospitals tied to gaps in quality is likely to yield much improved outcomes for patients. A measure of accountability for health plans helps them review trends over time and work with physicians and hospitals to improve the ways in which they engage patients using more optimal care management and care coordination (Cassel 2014). In addition, PAC measures could be used as a surrogate for quality in a consumer transparency tool to differentiate providers with regards to their performance.

Moreover, since these measures are claims based, there is minimal added burden for collecting the data, and it also avoids potential gaming that may occur for other measures that require reporting information to registries. Although use of administrative claims data in identifying conditions and measuring provider quality has been questioned, there are several studies in literature that acknowledge validity of its use (Normand 2007) (Quan 2009). Until more readily available data are at hand, use of administrative data to measure provider performance has steadily increased (Miller 2001), (NQF Quality Positioning System). Interestingly, in the current fee for service system, services for most PACs are rewarded by continued payment (except the CMS defined "never events") and hence to our advantage, adverse events surface in billing data. Claims based PAC measures; therefore serve as an alternative method to track adverse outcomes that do occur (Leibson 2008).

References:

1) deBrantes F, Rastogi A, and Painter M. "Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach." Health Serv Res 45.6.2 (2010 Dec): 1854-1871. doi: 10.1111/j.1475-6773.2010.01136x

2) Joynt KE, Gawande AA, Orav EJ, and Jha AK. "Contribution of Preventable Acute Care Spending to Total Spending for High-Cost Medicare Patients." JAMA 309.24 (2013): 2572-2578. doi: 10.1001/jama.2013.7103.

3) James JT. "A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care." J Patient Safety 9.3 (2013): 122-128.

4) See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

5) Yong, Pierre L., Robert Samuel Saunders, and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, D.C.: National Academies, 2010. Institute of Medicine of the National Academies, 17 Dec. 2010. Web.

6) BCBSNC: Blue Cross Blue Shield of North Carolina:
https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

7) Community Campaigns for Quality Care. "Recommendations to Reduce Potentially Avoidable Complications (PACs) among CalPERS Employees." Editorial. Calpers.ca.gov. Community Campaigns for Quality Care, June 2012. Web.

8) 2015 Bundled Payment Summit – Day 1, Track IV: Washington DC June 3-5.
http://www.bundledpaymentsummit.com/agenda/day1.html

9) Micaela P. McVary. “The Prometheus Model: Bringing Healthcare into the Next Decade.” Annals of Health Law Advance Directive 19 (2010): 274-284.

10) Colorado Business Group on Health: Healthcare Incentives Payment Pilot (HIPP): http://www.cbghealth.org/projects/reducing-costs/healthcare-incentives-payment-pilot-hipp/

11) Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. “An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care.” Health Aff (Millwood) 31.3 (2012): 560-8. doi: 10.1377/hlthaff.2011.1168.

12) Cassel, Christine, MD et al. "Getting More Performance from Performance Measurement." New England Journal of Medicine 371 (2014): 2145-147. Web.

13) Normand, Sharon-Lise T., Yun Wang, and Harlan M. Krumholz. "Assessing Surrogacy of Data Sources for Institutional Comparisons." Health Services and Outcomes Research Methodology Health Serv Outcomes Res Method 7.1-2 (2007): 79-96. Web.

14) Quan, H., N. Khan, B. R. Hemmelgarn, K. Tu, G. Chen, N. Campbell, M. D. Hill, W. A. Ghali, and F. A. Mcalister. "Validation of a Case Definition to Define Hypertension Using Administrative Data." Hypertension 54.6 (2009): 1423-428. Web.

15) Miller MR, Elixhauser A, Zhan C, and Meyer G. “Patient Safety Indicators: Using Administrative Data to Identify Potential Patient Safety Concerns.” Heath Services Research 36.6.2 (2001): 110-132.

16) NQF: Quality Positioning System ™. National Quality Forum, 2015. Web.: Available at http://bit.ly/1ijI5Ar, Last accessed June 29 2015.

17) Leibson CL1, Needleman J, Buerhaus P, Heit JA, Melton LJ 3rd, Naessens JM, Bailey KR, Petterson TM, Ransom JE, Harris MR. Identifying in-hospital venous thromboembolism (VTE): a comparison of claims-based approaches with the Rochester Epidemiology Project VTE cohort. Med Care. 2008 Feb;46(2):127-32. doi: 10.1097/MLR.0b013e3181589b92.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

The data included two years of administrative claims covering the period April 1, 2012 through December 17, 2014. There were a total 1,806 episodes of PCMDFR.

Because facilities with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. After this exclusion 22 (out of 380) facilities remained. Performance scores for these facilities are summarized in the following table:

Unadjusted PAC Rates:
- Median (IQR): 46.8% (39.5%, 55.2%)
- Range: 20% - 64.3%

Risk-Standardized PAC Rates (RSPR):
- Median (IQR): 46.2% (36.8%, 54.0%)
- Range: 20.8% - 62.5%

Please refer to the NQF_PCMDFR_all_codes_risk_adjustment 06.30.15.xls workbook under the “ProviderAttribution Reliability” tab to see specific results for each facility.

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.

Physician level of training and level of specialty certification have been shown to affect the risk of adverse events associated with ICD implant. An ICD Registry analysis found that physicians who implant more ICDs have lower rates of procedural complications and in hospital mortality (Krahn 2011) (Poole 2010).

A study analyzing data from the REPLACE registry reported a 4.0% complication rate in 1031 patients undergoing generator replacement and 15.3% in 713 patients with replacement and a lead addition. The REPLACE registry reported that ICDs were associated with a greater risk of complications (Poole 2010). Van Rees et al., who conducted a systematic review of major RCT's, concluded that both thoracotomy and non-thoracotomy ICD's had significantly higher in-hospital mortality and higher complication rates (Poole 2010). Mariana Parahuleva, in her extensive review of PCMDFR complications published as a book chapter states that, identification of factors contributing to complications may permit identification of high-risk individuals that warrant incremental monitoring and therapy to attenuate risk (Parahuleva 2011).

While PACs may not be completely eliminated, identifying their magnitude and understanding their causality, in particular for the most frequent or the most expensive, could lead to improving patient outcomes.

References:
1) Krahn AD, et al. "Predictors of Short-Term Complications After Implantable Cardioverter-Defibrillator Replacement: Results From the Ontario ICD Database". Arrhythmia Electrophysioly 4.2 (April 1 2011): 136-142.

2) Poole JE, et al. "Complication rates associated with pacemaker or implantable cardioverter- defibrillator generator replacements and upgrade procedures: results from the REPLACE registry." Circulation 122 (2010):1553–1561.

3) Parahuleva, Mariana. "8 Cardiovascular Implantable Cardioverter Defibrillator-Related Complications: From Implant to Removal or Replacement: A Review." Cardiac Defibrillation - Mechanisms, Challenges and Implications. Giessen, Germany: INTECH Open Access, 2011. Web.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Not applicable

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Frequently performed procedure, High resource use, Patient/societal consequences of poor quality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
The use of pacemakers / defibrillators (PCMDFR) has increased exponentially by a combination of increasing number of patients with cardiac disease, expanding indications of device use and improved technology (van Rees 2011) (Greenspon 2012). Publication of the second Multicenter Automated Defibrillator Implantation Trial (MADIT II) trial resulted in the Centers for Medicare and Medicaid Services (CMS) coverage of ICDs (implantable cardioverter defibrillator) for patients meeting the MADIT II inclusion and exclusion criteria (Reynolds 2003). There was a 55.6% increase in the use of PCMDFR's in the 17 years from1993 to 2009, with 2.9 million patients receiving permanent pacemakers. The economic impact of these devices is significant with hospital charges for ICD's increasing by 45.3% during the same period (Greenspon 2012).

PCMDFR use is associated with complications at all stages of device use, starting from implantation, in-hospital stay to post discharge (Bond 2012). Patients with any complication, compared with those with none, generated $7251 in increased hospital costs and 3.4

days in increased LOS after adjustment for baseline characteristics (Reynolds 2006). In a large study analyzing more than 30,000 Medicare beneficiaries undergoing PCMDFR implantations, 10.8% experienced one or more complications prior to discharge. The occurrence of a complication increased adjusted hospital costs by almost 20% with an incremental cost of complications exceeding $78million per 100,000 implanted defibrillators (Reynolds 2006).

In order to analyze and report ICD procedural complications, the CMS created the Medicare ICD registry in 2005, now maintained by the American college of Cardiology (ACC NCDR ICD Registry). An analysis of more than 350,000 ICD implantations included in the National Cardiovascular Data Registry–ICD Registry revealed 3.1% of patients experienced in hospital adverse events, 1.2% experienced major adverse events, and 0.4% died. Adverse events were lower (1.9%) with single-chamber ICD implants than with dual-chamber ICD implants (2.9%) or with biventricular ICD implants (4.1%). Specific adverse event rates included lead dislodgement (1%), hematoma (0.9%), pneumothorax (0.4%), and cardiac arrest (0.3%)(Freeman 2012).

Associated comorbidities, type of device used as well as physician volumes further influence the complication rates. Physician level of training and level of specialty certification have been shown to affect the risk of adverse events associated with ICD implant. Various studies have shown that the commonest complications like lead displacement, pneumothorax, infection and perforation are all PAC's and are influenced by physician experience and training (Bond 2012) (Reynolds 2006) (Grimm 2006) (Eberhardt 2005) (Parsonette 1998) (Johansen 2011). An ICD Registry analysis also confirmed this finding (Krahn 2011) (Poole 2010).

With the rise in use of pacemakers / defibrillators (PCMDFR) and the significant number of avoidable complications (Kirkfeldt 2013) (Koneru 2011) (Beyerbach 2014) that are under the provider's control, PAC measures in the pacemaker / defibrillator population is a high priority aspect of health care. The PAC measures look for all-cause harms in patients receiving these devices.

1c.4. Citations for data demonstrating high priority provided in 1a.3

1) van Rees, Johannes B., Mihály K. De Bie, Joep Thijssen, C. Jan Willem Borleffs, Martin J. Schalij, and Lieselot Van Erven. "Implantation-Related Complications of Implantable Cardioverter-Defibrillators and Cardiac Resynchronization Therapy Devices." Journal of the American College of Cardiology 58.10 (2011): 995-1000. Pubmed.gov. Web.

2) Greenspon, Arnold J., Jasmine Patel, Edmund Lau, Daniel Frisch, Reginald Ho, Behzad Pavri, Jorge Ochoa, and Steven Kurtz. "Trends In Permanent Pacemaker Implantation In The United States 1993-2009: Increasing Complexity Of Patients And Procedures." Journal of the American College of Cardiology 59.13 (2012). Web.

3) Reynolds, M. R. "MADIT II (Second Multicenter Automated Defibrillator Implantation Trial) Debate: Risk Stratification, Costs, and Public Policy." Circulation 108.15 (2003): 1779-783. Web.

4) Bond, Richard, Daniel Augustine, and Mark Dayer. "Pacemaker Complications in a District General Hospital." British Journal of Cardiology Br J Cardiol 19.2 (2012): 90-94. Web.

5) Reynolds, Matthew R., David J. Cohen, Aaron D. Kugelmass, Phillip P. Brown, Edmund R. Becker, Steven D. Culler, and April W. Simon. "The Frequency and Incremental Cost of Major Complications Among Medicare Beneficiaries Receiving Implantable Cardioverter-Defibrillators." Journal of the American College of Cardiology 47.12 (2006): 2493-497. Web.

6) "Hospital Registries." Quality Improvement for Institutions. American College of Cardiology, n.d. Web. 28 June 2015. <http://cvquality.acc.org/en/NCDR-Home/Registries/Hospital-Registries.aspx>.

7) Freeman JV, Y Wang, JP Curtis, PA Heidenreich, MA Hlatky. "Physician procedure volume and complications of cardioverter-defibrillator implantation." Circulation 125.1 (2012): 57-64.

8) Grimm, Wolfram, Belinda F. Flores, and Francis E. Marchlinski. "Complications of Implantable Cardioverter Defibrillator Therapy: Follow-Up of 241 Patients." Pacing and Clinical Electrophysiology Pacing Clin Electro 16.1 (1993): 218-22. Web.

9) Eberhardt, F., et al. "Long Term Complications in Single and Dual Chamber Pacing Are Influenced by Surgical Experience and Patient Morbidity." Heart 91.4 (2005): 500-06. NCBI. Web.

10) Parsonnet, Victor, Alan D. Bernstein, and Bruce Lindsay. "Pacemaker-implantation Complication Rates: An Analysis of Some Contributing Factors." Journal of the American College of Cardiology 13.4 (1989): 917-21. Web.

11) Johansen J, et al. "Infection after pacemaker implantation: infection rates and risk factors associated with infection in a

population-based cohort study of 46299 consecutive patients." Eur Heart J 32 (2011):991–8.

12) Krahn AD, et al. "Predictors of Short-Term Complications After Implantable Cardioverter-Defibrillator Replacement: Results From the Ontario ICD Database." Arrhythmia Electrophysioly 4.2 (2011): 136-142.

13) Poole JE, et al. "Complication rates associated with pacemaker or implantable cardioverter- defibrillator generator replacements and upgrade procedures: results from the REPLACE registry." Circulation 122 (2010):1553–1561.

14) Kirkfeldt, R. E., J. B. Johansen, E. A. Nohr, O. D. Jorgensen, and J. C. Nielsen. "Complications after Cardiac Implantable Electronic Device Implantations: An Analysis of a Complete, Nationwide Cohort in Denmark." European Heart Journal 35.18 (2013): 1186-194. Web.

15) Koneru, J. N., C. D. Swerdlow, M. A. Wood, and K. A. Ellenbogen. "Minimizing Inappropriate or "Unnecessary" Implantable Cardioverter-Defibrillator Shocks: Appropriate Programming." Circulation: Arrhythmia and Electrophysiology 4.5 (2011): 778-90. Web.

16) Beyerbach, Daniel M., MD, PhD, and Jeffrey N. Rottman, MD. "Pacemakers and Implantable Cardioverter-Defibrillators." Pacemakers and Implantable Cardioverter-Defibrillators. Medscape, 2 May 2014. Web. <http://emedicine.medscape.com/article/162245-overview>.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable

1d. Composite Quality Construct and Rationale

1d.1. **A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
  - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient); or
  - any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient).

1d.1. Please identify the composite measure construction: any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

1d.2. **Describe the quality construct, including:**

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

We classify PACs into two types: Type 1 PACs are directly related to the index condition and are often controlled by the servicing provider; Type 2 PACs, on the other hand result from patient safety failures and could be reduced by better systems and better processes in care. Both types of PACs could occur in any setting and so could be identified through any type of claims coming in the administrative dataset, including in-patient, out-patient, or professional claims. PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1.

The PAC measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs), Hospital Inpatient Quality Reporting measures, Avoidable Readmissions, AHRQ defined patient safety indicators (PSIs), NQF endorsed patient safety measures

such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates.

All defined PACs, irrespective of their type, or site of occurrence, are aggregated to create an overall comprehensive, composite measure. They all have equal weighting, since they are measured simply by the frequency of their occurrence.

1d.3. **Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.**
Each individual PAC, when measured in isolation, provides a very limited picture of the performance of the provider(s) who are managing or co-managing the care of the patient. However, looking at all the PACs that may occur individually or concurrently in a patient with a given episode provides a comprehensive picture of the care received by the patient for that particular condition or illness.

Additionally, the frequency of occurrence of individual PACs may be so low that it may require very high sample sizes from individual providers to achieve any meaningful and reliable comparisons. But aggregating all the PACs into a single quality metric creates meaningful scores that can be compared across providers even with relatively smaller sample sizes.

Additionally, a comprehensive measure is easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

1d.4. **Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**
In constructing the comprehensive composite PAC measure, each component PAC, as clinically defined by the subject matter experts, was given the same weight so that arbitrary weights may not bias the results. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence. As such, the patient is the ultimate unit of measurement and if the patient incurred any PAC during the episode, then that counts against the numerator.

Since the emphasis of the PAC measure was to simply identify the occurrence of PACs in any setting, aggregation of the PAC counts to create a comprehensive quality score with equal weights has been met with overall support from the clinical working groups as well as from the implementation sites.

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 subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

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

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

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Care Coordination : Readmissions, Safety, Safety : Complications, Safety : Healthcare Associated Infections, Safety : Medication Safety, Safety : Readmissions, Safety : Venous Thromboembolism

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.)*
http://www.hci3.org/ecr_descriptions/ecr_description.php?version=5.2.006&name=PCMDFR&submit=Submit

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_PCMDFR_all_codes_risk_adjustment_06.30.15-635719851913348171.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of patients who underwent a pacemaker/defibrillator implantation (PCMDFR), are followed for at least 30-days, and have one or more potentially avoidable complications (PACs) during the episode time window.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The time window includes a 7-day look-back period and a 30-day look-forward period from the PCMDFR trigger claim.

S.6. **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, 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.
Patients that have triggered a PCMDFR episode, are followed for at least 30-days, and are identified as having services for potentially avoidable complications (PACs). PACs may occur during the index stay or during the 30-day post discharge period. The enclosed excel workbook entitled NQF_PCMDFR_all_codes_risk_adjustment_06.30.15 gives the detailed codes for PACs in the tabs entitled PACs I-9 and PACs I-10.

Services for PACs are identified as follows:
a. Any Index stay that has a PAC diagnosis code in any position except in the PRIMARY (principal) position is considered as having a potentially avoidable complication
b. Any readmission to an acute care facility 2 days or later after discharge but within 30-days post-discharge, that is relevant to PCMDFR
c. Any admission to a post-acute care facility, that is relevant to PCMDFR and has a PAC code in any position on the claim
d. Any other service (professional, outpatient facility, ancillary) that is relevant to PCMDFR and has a PAC code in any position on the claim

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Adult patients aged 18 years and above who underwent a Pacemaker/defibrillator implantation - PCMDFR) procedure and are followed for at least 30-days.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions, Populations at Risk : Veterans, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Please refer to the enclosed excel workbook entitled
NQF_PCMDFR_all_codes_risk_adjustment 06.30.15

The target population is identified using the following criteria:
1. Using administrative claims database, patients undergoing PCMDFR are identified using one of the following criteria:
 a. Patients with a procedure code of PCMDFR in any position on an in-hospital stay claim with a qualifying diagnosis code relevant to the PCMDFR procedure.
 b. Patients with a procedure trigger code of PCMDFR in any position on an outpatient facility claim with a qualifying diagnosis code relevant to the PCMDFR procedure.
 c. Patients having a professional service carrying a trigger code of PCMDFR in any position.
The trigger codes for PCMDFR and the qualifying diagnosis codes are provided in the tab called "Triggers I-9" or "Triggers I-10".

2. The patient should have continuous enrollment for the entire time window with no enrollment gap, with the entity providing the data (so we can ensure that the database has captured most of the claims for the patient during the episode time window).
3. The patient should have a complete episode time window in the claims data – so the end date of the episode should not be past the database claims end date.
4. Patient should be at least 18 years of age
5. Patients that have a trigger code on a professional claim and have no associated facility bill are considered as having an orphan (incomplete) episode and are dropped from analysis.

Once the episode is triggered all relevant claims are assigned to the episode. Relevant claims could be inpatient facility claims, outpatient facility claims, professional services, laboratory services, imaging services, ancillary claims, home health, durable medical equipment as well as pharmacy claims across the entire continuum of care centered around the patient's episode of care. Relevant claims are identified as those that have a diagnosis code that matches the codes in the typical Dx codes tabs (Typical Dx I-9 or Typical Dx I-10), or in the PAC Dx codes tab (PACs I-9 or PACs I-10) AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant readmissions and relevant admissions to post-acute care facilities are also included in the denominator.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Denominator exclusions include exclusions of either "patients" or "claims" based on the following criteria:
1. "Patients" excluded are those that do not meet the enrollment criteria. If patient has an enrollment gap for any time period during the episode time window, it is considered as an enrollment gap
2. "Patients" are also excluded if the cost of the episode is an outlier at greater than 99th percentile or less than 1st percentile value for all episodes. This is another way to ensure that episodes are complete as well as they do not bring in random noise into the analysis due to inappropriate codes or services.
3. "Claims" are excluded from the PCMDFR measure if they are considered not relevant to PCMDFR care.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
Denominator exclusions include exclusions of "patients" as well as "claims" not relevant to PCMDFR care. Please refer to the enclosed excel workbook entitled (NQF_PCMDFR_all_codes_risk_adjustment 06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
 a. If age is < 18 years
 b. If gender is missing
 c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
 d. If the episode time window extends beyond the dataset end date (this helps eliminate incomplete episodes).
 e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure based on the following criteria:
 a. If none of the diagnosis codes on the claim are on the list of relevant diagnosis codes (either typical Dx or PAC Dx) for PCMDFR.
 b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for PCMDFR.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
None

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Conceptual Model
Variations in outcomes across populations may be due to patient-related factors or due to provider-controlled factors. When we adjust for patient-related factors, the remaining variance in PACs are due to factors that could be controlled by all providers that are managing or co-managing the patient, both during and after hospitalization.

Statistical Method:
Logistic Regression model to determine the probability of a patient incurring a PAC
Demographic variables, comorbid conditions, as well as clinical severity indicators are fed as independent risk factors into the model. Risk Factors are collected historically. Subtype information is collected from the index claim and any look-back period, if relevant. Subtypes are clinical severity indicators suggesting severity of the episode itself, for example, diagnosis of cardiomyopathy in a PCMDFR patient. For each patient the "predicted" coefficients from the risk adjustment models are summed to give the predicted probabilities of the occurrence of a PAC.

Risk Factors :(Please refer to the enclosed excel workbook entitled (NQF_PCMDFR_all_codes_risk_adjustment 06.30.15.xls). The risk factors along with their codes are listed in the tabs called "All Risk Factors I-9" and "All Risk Factors I-10" and also listed below:

AGE CONTINUOUS VARIABLE
GENDER FEMALE = 1 (MALE IS REFERENCE = 0)

Risk Factor #	Risk Factor Name
RF0101	Anoxic Brain Damage, persistent vegetative state
RF0102	Delirium, Meningitis, Encephalitis
RF0103	Previous Stroke, Paralysis
RF0104	Cerebral Palsy and Other Paralytic Syndromes
RF0105	Spinal Cord Disorders/Injuries
RF0106	Polyneuropathy
RF0107	Multiple Sclerosis
RF0108	Convulsions, Epilepsy
RF0109	Dementia
RF0110	Parkinson´s and Huntington´s Diseases
RF0111	Cerebrovascular Disease
RF0115	after care, rehabilitation
RF0201	visual loss, blindness, retinal tear, detachment
RF0301	ENT, Upper Respiratory Problems
RF0401	Respiratory Failure, O2, ventilator dependence
RF0402	Advanced COPD, Asthma
RF0403	Empyema, bronchiectasis, Pneumonias
RF0404	Aspiration Pneumonia, Laryngeal Problems
RF0406	TB, Pneumoconiosis, Aspergillosis
RF0407	Tobacco use, Lung disease due to External Fumes
RF0408	Other Lung Disease
RF0501	Previous Shock, Syncope, Vent Fibrillation
RF0503	Advanced CHF
RF0504	Cardiomyopathy, valve disorders
RF0505	Cardiac Arrhythmias, Heart Block
RF0506	Pacemaker, AICD
RF0507	Endocarditis, Other post surgical cardiac problems
RF0508	Other Cardiovascular Disease
RF0511	DVT, Pulm Embolism, Pulm Heart Disease

RF0512 Unstable Angina
RF0513 Hypotension, chronic, orthostatic
RF0514 Hyperlipidemia
RF0515 Intraaortic Balloon Pump
RF0516 ventricular assist device, ecmo, prolonged bypass
RF0517 Previous electrophysiology studies, cryoablation
RF0518 Recent AMI
RF0519 Previous PCMDFR
RF0520 Previous CABG
RF0521 Previous Heart & Valve Surgery
RF0522 Previous aortic reconstruction
RF0523 Previos carotid endarterectomy
RF0524 Aortic and peripheral vascular disease
RF0525 Advanced Aortic and Vascular Disease
RF0601 GI Bleed
RF0602 Intestinal Obstruction/Perforation
RF0603 Acute Gastritis, Duodenitis
RF0604 Gastroduodenal Ulcer
RF0606 Intestinal Uro-genital Fistula
RF0607 Abdominal hernia w complications
RF0608 Vascular insufficiency of intestine
RF0609 Inflammatory Bowel Disease
RF0610 Irritable Bowel
RF0611 Diverticulitis, Meckel´s
RF0612 Digestive congenital anomalies
RF0613 Intestinal infection
RF0614 Esophageal Perforation, Hmg, Barretts, Compl Hiatal Hernia
RF0615 Abnormal weight loss
RF0616 Achalasia, Esophageal spasm, Stricture, Dysphagia
RF0617 GERD, Hiatal Hernia, Other Upper GI Disorders
RF0618 Previous Bariatric Surgery
RF0619 Hx of colon polyps, family Hx of colon cancer
RF0620 Enterostomy, GI devices, lap band
RF0701 Pancreatic Disease
RF0702 Perforation, fistula GB, bile duct, pancreas
RF0703 Gall stones, cholecystitis
RF0704 End-Stage Liver Disease
RF0705 Hepatitis, Cirrhosis, Other Hepatbiliary Disorders
RF0706 Recent Gall Bladder, Hepatobilary Surgery
RF0707 Acute Pancreatitis, pseudo cyst
RF0801 Bone/Joint/Muscle Infections/Necrosis
RF0802 Muscular Dystrophy
RF0803 Osteoporosis, ostetits deformans, pathological fracture
RF0804 Rheumatoid Arthritis and Inflammatory Connective Tissue Disease
RF0805 Gout and other crystal arthropathies
RF0806 Other arthropathies
RF0807 Osteoarthritis
RF0808 Joint Deformities
RF0809 Knee derangements
RF0810 Traumatic Dislocation Knee
RF0811 Dislocation Hip
RF0812 Synovitis, Ruture Tendon
RF0813 Status Knee Replacement
RF0814 Status Total Hip Replacement
RF0901 Decubitus Ulcer
RF0902 Skin and wound problems
RF1001 Diabetes, poor control

RF1002 Advanced diabetes
RF1003 diabetes
RF1101 Acute renal failure
RF1102 Dialysis Dependent
RF1103 Nephritis
RF1104 Chronic renal failure
RF1105 Urinary Tract Infections
RF1301 Endometriosis
RF1302 Fibroid uterus, benign tumors of female organs
RF1303 Pelvic Inflammatory disease
RF1304 Uterine prolapse, cystocele, vaginocele
RF1305 Female Harmonal Disorders
RF1306 Ovarian, Broad Ligament Disorders
RF1308 Other disorders of uterus, cervix
RF1309 Menopausal Disorders
RF1310 Menstrual Disorders
RF1401 Multiparity, multigravida
RF1402 Elderly Primi, other
RF1403 Poor obstetric history
RF1406 Cervical incompetence
RF1407 Abnormalities of uterus, female genital tract
RF1408 Hypertension, pre-eclampsia in Pregnancy
RF1409 Severe pre-eclampsia w HTN, Eclampsia
RF1410 Maternal, gestational diabetes, large for date
RF1411 Genital Herpes
RF1412 Infections of genitourinary tract, venereal disease in pregnancy
RF1413 Infectious Diseases in Mother
RF1414 Cardiovascular disease in Mother
RF1415 Mental Disorders in Mother
RF1416 Epilepsy in Mother
RF1417 Liver and biliary tract disorders in mother
RF1418 Kidney Disease in Mother
RF1419 Other Maternal conditions
RF1421 Cephalopelvic Disproportion due to maternal causes
RF1436 Peripartum Cardiomyopathy
RF1441 Previous Cesarean section
RF1450 Maternal Obesity, previous Bariatric Surgery
RF1454 Previous Rupture Uterus, Obstetrical Trauma
RF1458 Complicated Pregnancy Delivery
RF1460 Thrombophlebitis, DVT during Pregnancy
RF1461 Puerperal Sepsis, other major puerperal complications
RF1462 Obstetrical Embolism, Air, Amniotic Fluid, Pulm, Pyemic
RF1467 Tobacco Use in Mother
RF1601 Bleeding Disorders
RF1602 Severe Hematological Disorders
RF1603 Disorders of Immunity
RF1604 Nutritional and other Anemias
RF1605 Long-term use of anticoag, Aspirin
RF1701 Head and Neck Cancers
RF1702 Lung and Intrathoracic Cancers
RF1703 Neuroendocrine, Myeloproliferative Cancers
RF1704 Poorly differentiated, Secondary, Metastatic Cancers
RF1705 Other Tumors
RF1706 Acute Leukemia
RF1707 Cancer uterus, localized female organs
RF1708 Colorectal, Hepatobiliary and other GI cancers
RF1709 Breast, Prostate, Thyroid cancers

RF1710 Testicular Cancer and localized of male organs
RF1711 Cancer of Bladder and Urinary Tract
RF1712 Musculoskeletal Cancers
RF1801 Sepsis, MRSA, Opportunitistic infections
RF1901 Schizophrenia
RF1902 Major Depressive, Bipolar, and Paranoid Disorders
RF2001 Drug/Alcohol Psychosis
RF2002 Drug/Alcohol Dependence
RF2101 Drug Reactions, long term use of drugs
RF2102 Intra-abdominal injury
RF2201 Extensive Third-Degree Burns
RF2301 Major Organ Transplant Status
RF2302 Artificial Openings for Feeding or Elimination
RF2303 Complications of Medical & Surgical Care and Trauma
RF2304 severe morbid obesity
RF2305 morbid obesity
RF2306 obesity
RF2307 mild sleep apnea, hypoventilation
RF2308 moderate sleep apnea, hypoventilation
RF2309 obstructive sleep apnea
RF2310 Severe Protein-Calorie Malnutrition
RF2311 Mild-mod malnutrition
RF2401 Severe Head Injury
RF2402 Major Head Injury
RF2403 Vertebral Fractures without Spinal Cord Injury
RF2404 Falls, Fractures
RF2405 Amputation
RF2501 HIV/AIDS

Subtypes for PCMDFR
Insertion of Pacemaker
Insertion of leads
Insertion of Generator alone
Insertion of Defibrillator
Pacemaker, Defibrillator in place
Malfunction / Complication of Heart Device, H
Highgrade Heart Block
Other Heart Blocks / Conduction Disorders
History of Sudden Death
Ventricular Arrhythmias
Sinus Node Dysfunction
Supraventricular Tachyarrhythmias
Other cardiac arrhythmias
Atrial Flutter / Fibrillation
Heart Failure, Cardiomegaly
Cardiomyopathy
Heart Valve Disorders
Heart Aneurysm and other Sequelae of AMI
Shock / Cardiac Arrest
Unstable angina
STEMI
Subendocardial infarct
Previous CABG, PCI
Recent AMI
Acute CHF / pulm edema
Diastolic Heart Failure
Previous heart valve replacement

Pulmonary heart disease
Congential Heart Disease w Structural Defects
Coronary Artery Anomaly
Hypertensive Heart Disease
Hypertensive Heart Disease w Heart Failure
Hypertensive Heart Disease w Heart Failure & CKD
Renovascular and other secondary hypertension
Acute pericarditis
Myocarditis
Chronic, adhesive, constrictive pericarditis
Other pericarditis
Other heart disease
Transplanted Heart
Tumor of the Heart
Aortic and peripheral vascular disease
Artificial Heart, Assist Device
Severe Morbid Obesity
Morbid Obesity
Obesity
Overweight
Obstructive sleep apnea
Sleep Apnea

The prevalence of the risk factors in our reference dataset are listed in the enclosed workbook entitled NQF_PCMDFR_all_codes_risk_adjustment 06.30.15.xls – see tab "Risk Factor Prevalence". The output of the regression model are given in the same workbook in the tab "Risk Model'.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
Please refer to the enclosed excel workbook entitled (NQF_PCMDFR_all_codes_risk_adjustment 06.30.15.xls).

Assembling the Denominator:

Using administrative claims database, patients undergoing a PCMDFR are identified using one of the following criteria: 1) Patients with a procedure code of PCMDFR in any position on an in-patient of an out-patient facility claim with a qualifying diagnosis code relevant to the PCMDFR procedure, 2) Patients having a professional service carrying a trigger code of PCMDFR in any position. The trigger codes for PCMDFR are provided in the tab called "Triggers I-9" or "Triggers I-10".

Patients are retained if they are 18 years of age or more, do not have a missing gender, have a complete episode time window in the database, have a maximum of 30-dayno enrollment gap for the entire episode time window, and have no outlier episode costs. All relevant professional, laboratory, imaging, ancillary and other claims that are incurred during the episode time window are included as part of the episode. Claims are considered relevant to PCMDFR care if they have one of the diagnosis codes, as listed on the tab entitled Triggers I-9, Triggers 1-10, PACs I-9, PACs I-10, Typical Dx I-9, or Typical Dx I-10 in any position on the claim AND a procedure code as identified in the Relevant Procedures I-9 & I-10 tab in the enclosed workbook. Relevant readmissions and relevant admissions to post-acute care facilities are also included in the denominator. All relevant pharmacy claims carrying codes that match the ingredients listed in the Pharmacy tab of the enclosed workbook are also included as part of the episode.

If a patient has more than one concurrent episode, and the claim is relevant to both episodes, the claim could get multi-assigned, except in the case of procedural episodes that get carved out with respect to the index stay. So if an inpatient stay claim carried a procedure code that matched the trigger procedure code for PCMDFR but they also had a qualifying diagnosis code for CAD (coronary artery disease), the stay claim would trigger both episodes concurrently, but get uniquely assigned to PCMDFR and not be counted with CAD.

Once all the episodes are assembled, episodes that match the exclusion criteria, such as those with outlier costs, are flagged (those with total episode costs less than 1st percentile or greater than 99th percentile), and excluded from the final analysis.

Assembling the Numerator:

For every episode included in the denominator, services are flagged as having a PAC (potentially avoidable complication) based on the criteria listed below:

Any Index stay that has a PAC diagnosis code in any position except in the PRIMARY (principal) position is considered as having a potentially avoidable complication

Any readmission to an acute care facility 2 days or later after discharge but within 30-days post-discharge, that is relevant to PCMDFR

Any admission to a post-acute care facility, that is relevant to PCMDFR and has a PAC code in any position on the claim

Any other service (professional, outpatient facility, laboratory, imaging, ancillary) that is relevant to PCMDFR and has a PAC code in any position on the claim

Relevant claims that do not have any PAC codes, and do not qualify as a PAC based on the criteria outlined above, are listed as typical claims. All included relevant pharmacy services are flagged as typical. Patients that have even a single PAC claim are counted as part of the numerator.

Calculating the measure:

Proportion of PCMDFR patients that have PACs, is simply the ratio of patients with PACs within the PCMDFR population and is called the PAC rate as shown in the equation below:

PAC rate = Patients with PCMDFR that have at least one PAC claim / Total number of PCMDFR patients

A flow chart demonstrating the series of steps and the counts of patients at each step is shown in tab entitled Decision Tree of the enclosed workbook called NQF_PCMDFR_all_codes_risk_adjustment 06.30.15.xls

Drill Down Calculations:

Further analysis from this construct helps create actionable reports.

For example as shown in the tab labeled PAC overview, not only do we have the PAC rate for a population, we can break them down by the PAC type – type 1 being directly related to PCMDFR and so actionable by the servicing physician, while type 2 PACs are related to patient safety failures and can be improved by process improvement. Additionally, analyzing what portion of the PACs occur during the index stay, vs. in the post-discharge period and how many are due to readmissions helps focus strategies in reducing them.

Risk Adjustment:

Once we have the observed PAC rates, we risk-adjust them for patient factors such as patient demographics, comorbidities collected historically, and for severity of illness or procedure using subtypes collected from the index stay and / or look-back period. This

helps adjust for factors outside the providers control and levels the playing field for provider performance comparisons.

Unit of Analysis:
The unit of analysis is the individual episode.

Dependent Variable:
The dependent variable is a dichotomous variable indicating whether an episode had one or more claims assigned as a PAC (=1) or not (=0).

Independent Variables:
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_PCMDFR_all_codes_risk_adjustment 06.30.15.xls

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., unstable anginacardiomyopathy). Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_PCMDFR_all_codes_risk_adjustment 06.30.15.xls

As mentioned previously, to avoid creating perverse incentives all comorbidities and subtypes are identified prior to or at the very start of the episode. None are identified during the episode period.

Statistical Methods
We use logistic regression to model the probability of at least one PAC occurring during the episode. Only comorbidities and subtypes are included in the models as covariates if they are present in at least 10 episodes to prevent unstable coefficients. No further model building is conducted after the initial models are built. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but it does not make it a priority that all covariates in the model be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity of a episode condition, and lets each regression model determine for itself which of the factors are more significant for a specific episode. Non-significant covariates in episode models can not overly influence predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

When more than one line of business is included in the data, separate models are calculated for each sample (i.e., commercial, Medicaid etc.).

Provider Attribution and calculating PAC rates by provider:

Once episodes are constructed they are attributed to providers based on one of the various attribution rules. For PCMDFR, episodes are attributed to the facility where the episode triggered, or, if the episode is triggered off a professional claim, it is attributed to the first facility claim that overlaps the professional trigger claim date.episodes are attributed to the facility where the episode triggered, or, if the episode is triggered off a professional claim, the facility listed on the first facility claim that occurs during the episode.

Using the logistic regression technique described above, a model is developed that gives estimates for each risk factor and subtype for the patients in the population analyzed. These estimates are used to develop patient-level probabilities for the occurrence of PACs. The patient-level probability estimates are summed to construct aggregated measures (e.g., facility/provider-level). This method is similar to the methods employed by the Centers for Medicare and Medicaid Services (CMS) and endorsed by the National Quality Forum (NQF) to construct similar facility- and practice-level measures (i.e., mortality, readmissions, etc.):
1. For each provider, the number of actual observed occurrences of the outcome is summed across all attributed patients with that episode, to give the observed PAC rates for the provider.
2. Similarly adjusted probabilities from the risk adjustment models are summed across all attributed patients to give expected PACs

for the provider.
3. The observed sum is then divided by the summed probabilities (O/E). This number yields whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). This calculation yields a practice-level unstandardized performance ratio.
4. To facilitate accurate comparisons of rates across units of analysis, this ratio is then standardized to the community rate using the indirect method. Specifically, the provider-level rate is multiplied by the expected community rate, calculated as the sum of adjusted probabilities for every individual in the sample across all providers in the analysis. This measure, known as the standardized rate, represents what the unit's risk-adjusted rate would be for the outcome of interest if its patient population was reflective of the of the overall community.

The formula for this calculation is as follows:

Adj Outcome_j={(SUM Observed_ij)/(SUM Prob_ij)} × {(SUM Prob_i) / (# of episodes)}
Where individual is attributed to unit of analysis j (e.g., practice, provider, etc.)

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset will apply to another.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
If patient related data is missing, the case is deleted from both the numerator and the denominator

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The information is based on a two-year claims database from a large regional commercial insurer. The database has over 3.2 million covered lives and $25.9 billion in "allowed amounts" for claims costs. The database is an administrative claims database with medical as well as pharmacy claims.

The methodology can be used on any claims database with at least two years of data and a minimum of 150 patients with the index condition or hospitalization. Having pharmacy data adds to the richness of the risk-adjustment models.
The calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.
We also plan on providing a limited automated analysis, at no cost, on our website.
The methodology has been tested on databases of several health plans as well as on a few employer databases.

No data collection instrument was used.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Team, Facility, Health Plan, Population : National, Population : Regional, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Ambulatory Surgery Center (ASC), Hospital/Acute Care Facility, Other
If other: Across the care continuum

S.28. **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.)* The individual complications are considered measurable components. Separate specifications are not required for this measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2752_PCMDFR_Testing_Reliability_Validity_HCI3.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Composite Measure Title: 2752
Measure Title: Proportion of Patients undergoing Pacemaker / Defibrillator Implantation (PCMDFR) that have a Potentially Avoidable Complication (during the episode time window)
Date of Submission: 06/30/15
Composite Construction:
☐Two or more individual performance measure scores combined into one score
☐ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)
☐ Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specifications** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 pages (*including 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.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.
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. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. 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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

The information is based on a two-year administrative claims database from a large regional commercial insurer. The database contains medical and pharmacy claims on over 3.2 million covered lives and more than $25.9 billion in "allowed amounts" for costs.

1.3. What are the dates of the data used in testing? April 1, 2012 – December 17, 2014

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Integrated Delivery System	☐ other:

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

There were a total of 380 facilities in the data set. Because providers or facilities with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the reliability calculations. After this exclusion, there were 22 facilities left.

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

After exclusions (see 2b.3.1 below), there were a total of 1,806 episodes of PCMDFR were included in the testing and analysis. Patients in these episodes were, on average, 54.1 years of age (range 18-64) and 31% were female. We did not have race information on these patients. All patients for this analysis had a trigger inpatient claim of PCMDFR as identified in our code tables.

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.

For the reliability analysis, we restricted the data to only facilities with at least 10 attributed episodes. For risk adjustment, all episodes were used in the analysis, regardless of the facility to which they were attributed.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
None of the analyses included SDS variables.

_____

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*)

We assessed the reliability of the measure to demonstrate that it sufficiently differentiates performance between providers using the beta-binomial method, which is applicable to measures of this type. Reliability is a measure that distinguishes the signal (the extent of performance variation between entities that is due to true differences in performance) from statistical noise. Our approach follows directly from the methods outlined in the technical report "The Reliability of Provider Profiling: A Tutorial" by J.L. Adams.

Reference:

Adams JL. The Reliability of Provider Profiling: A Tutorial. Rand Corporation. http://www.rand.org/pubs/technical_reports/TR653.html.

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*)

The beta-binomial failed to produce statistically significant parameters. Therefore the reliability scores calculated show low reliability and we will be unable to differentiate provider performance in this dataset. For detailed calculations, refer to the workbook entitled, NQF_PCMDRF_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab to see provider-specific results.

As an ad hoc analysis, we tested the measure in another commercial dataset, and found decent reliability scores and providers with sample sizes as low as 22 patients, had a reliability >0.7. Results of this analysis are also displayed in the enclosed workbook entitled, NQF_PCMDRF_all_codes_risk_adjustment_06.30.15.xls, under the "ProviderAttribution Reliability" tab.
.

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?*)

Reliability scores can vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or variation across patients within providers) whereas a reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

There is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (see Adams, 2009 cited above).

In this dataset, based on our analysis, we were unable to report adequate reliability scores, suggesting that statistically the measure may not adequately differentiate performance between facilities.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

_____

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Content validity was built into the development of the definitions of potentially avoidable complications (PACs). This involved working with clinicians who are experts in their respective fields and specific to the episodes for which PACs are being measured. In particular, the clinical experts focused on whether or not a potentially avoidable complication can be deemed as such for a specific episode of care, and help defined and review all of the diagnosis and procedure codes for each PAC. The enclosed link lists clinicians who have participated in the various Clinical Working Groups (http://www.hci3.org/content/clinical-working-group-contributors). Some of the clinical experts have also participated in monthly webinars that highlight the clinical aspects of these measures (http://www.hci3.org/content/using-ecrs-providers).

In addition, we illustrate that our measure has face validity in several ways.

Beyond the up front work performed by clinical experts, the validity of the measure has also been tested in various real world settings. For example, we have presented results of claims data analyses that reveal the frequency and costs of PACs to physicians in several different healthcare systems involved in our pilot site implementations, as well as to medical directors from the employer coalitions and the health plans that provided the dataset to run the analyses. Some of these implementations include the Pennsylvania Employee Benefits Trust Fund and local provider groups and hospital, Horizon Blue Cross Blue Shield of NJ and many

physicians and health systems.

In addition, we have performed dozens of analyses of very large claims data sets and reported results of rates and costs of PACs to policy makers, health plan leaders and physician leaders from different states. These include:

- Vermont Payment Reform Commission
- Maine Health Management Coalition
- WellPoint / Anthem CT
- NY State Medicaid
- CT Medicaid
- CO All-payer Claims Database, Center for Improving Value in Health Care

These analyses and their results have influenced, and continue to influence, the development of various public reporting, payment reform and delivery system reform efforts. To-date, we have never experienced either wholesale or partial rejection of the results of analyses showing rates of PACs, which demonstrates the level of acceptability – face validity – of the measures from the payer, policymaker, employer and payer communities.

As importantly, measures of potentially avoidable complications have face-validity with consumers. In a series of focus groups, Judy Hibbard and colleagues[1] examined the impact of presenting information about price and quality of certain providers in influencing the decisions of consumers. They tested the validity of PACs as a discriminator of quality, as well as other measures of quality, and used the dollar symbol to illustrate the level of price, much like is done for restaurant reviews. When the PAC measure was used, respondents selected the providers with the lowest PAC rates with a high level of confidence in choice, and used it as a surrogate for a strong quality signal. To the contrary, when more standard measures of quality were used, consumers tended to ignore them and use price as a surrogate for quality. As such, what the researchers found is that the very framing of potentially avoidable complications as an indicator of potential harm, is an effective way of communicating the quality of care. And when measures of PACs were presented in conjunction with price, consumers intuitively accepted the logical relationship between low PACs – fewer "defects" – and lower price.

Finally, our measure definitions encompass several other measures that are accepted as being valid complications of care and are widely used throughout the country. These include CMS defined Hospital Acquired Conditions (HACs)[2], Hospital Inpatient Quality Reporting measures [3], Avoidable Readmissions [4,5], AHRQ defined patient safety indicators (PSIs) [6], NQF endorsed patient safety measures such as patient fall rates, pressure ulcer rates, and peri-operative pulmonary embolism or deep vein thrombosis rates [7].

References:

[1] Hibbard JH, Greene J, Sofaer S, Fiminger K, and Hirsh J. An Experiment shows that a well-designed report on Costs and Quality can help consumers choose High-Value Health Care. *Health Affairs* 2012; 31(3): 560-568. doi: 10.1377/hlthaff.2011.1168

[2] CMS defined Hospital Acquired Conditions: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Hospital-Acquired_Conditions.html

[3] CMS operated Hospital Inpatient Quality Reporting Program: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

[4] Jencks SF, Williams MV, and Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. N Engl J Med 2009 (Apr); 360 (14): 1418-1428. doi: 10.1056/NEJMsa0803563.

[5] Casalino LP, Pesko MF, Ryan AM et.al. Small Primary Care Physician Practices have low rates of Preventable Hospital Admissions. Health Affairs, 2014; 33(9): 1-9. doi: 10.1377/hlthaff.2014.0434.

[6] Agency of Healthcare and Quality defined Patient Safety indicators: http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx

[7] NQF endorsed measures: Quality Positioning System: http://bit.ly/1E5ZdP7

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Not applicable.

2b2.4. 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?*)

Given the significant clinical input that went into developing the measure, the widespread use and acceptance the measure has gained among a wide variety of individuals and organizations across the health system (public and private payers, clinicians, consultants, patients, etc.) [1-13], and the parallels between this measure and other measures that are in widespread use, this demonstrates that the measure has strong face validity.

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)

2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.

3. de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.

4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.

5. Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.

6. Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.

7. François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)

8. de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

9. Rastogi A, Mohr BA, Williams JO, Soobader MJ, de Brantes F. Prometheus Payment Model: Application to Hip and Knee Replacement Surgery. Clin Orthop Relat Res 2009; 467(10): 2587-2597.

10. de Brantes F and Rastogi A. Evidence-Informed Case Rates: Paying for Safer, More Reliable Care. The Commonwealth Fund 40, publ. 2008; 1146:1-14.

11. de Brantes F, Gosfield A, Emery D, Rastogi A and G. D'Andrea, "Sustaining the Medical Home: How Prometheus Payment Can Revitalize Primary Care", Robert Wood Johnson Foundation Report, May 2009, http://www.rwjf.org/pr/product.jsp?id=42555, accessed October 2009.

12. de Brantes F, Camillus J. Evidence-informed case rates: a new health care payment model [Internet]. New York (NY): Commonwealth Fund; 2007 Apr [cited 2007 May 20]. Available from: http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=478278, Accessed Aug 1 2013.

13. Satin DJ, and Miles J. Performance Based Bundled Payments: Potential Benefits and Burdens. Available from: http://student.med.umn.edu/p4p-new/sites/default/files/MN%20Med%20Bundles%20Special%20Report%20-%20Satin.pdf, Accessed Aug 1 2013.

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

No formal exclusion testing was done since no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers.

Exclusions included exclusions of "patients" as well as "claims" not relevant to PCMDFR care. Please refer to the enclosed excel workbook entitled (NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls)

1. "Patients" are excluded from the measure if they meet one of the following criteria:
   a. If age is < 18 years
   b. If gender is missing
   c. If they do not have continuous enrollment for the entire time window with a maximum of 30 day enrollment gap with the entity providing the data (this helps determine if the database has captured most of the claims for the patient in the time window).
   d. If the episode time window extends beyond the dataset end date (this helps eliminate incomplete episodes).
   e. The episode cost is an outlier (less than 1st percentile or greater than 99th percentile value for all episodes of the same type). This eliminates extreme variation that may result from random outlier events.

2. "Claims" are excluded from the measure based on the following criteria:

a. If none of the diagnosis codes on the claim are on the list of relevant diagnosis codes (either typical Dx or PAC Dx) for PCMDFR.
b. If none of the procedure / CPT codes on the claim are on the list of relevant procedure codes for PCMDFR.

2b3.2. 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*)

We started with a total PCMDFR population of 3,968 episodes. After all the exclusions were applied, the remaining PCMDFR population included in the analysis consisted of 1,806 episodes. As mentioned above, no real exclusions were done. The only patients excluded were the ones that had incomplete or missing data and those that would not have given a homogenous population such as outliers. As such, no formal exclusion testing was done.

2b3.3. 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*)

No formal analysis was done on the impact of exclusions on performance scores.

Descriptive Explanation:

Exclusions of patients were for the following reasons. Some are for comparative purposes and some for medical reasons.

(a) Comparative Purposes:
We excluded patients that did not have complete enrollment for the entire episode time window. This was done to ensure that the database had complete information on patients to be able to create the entire episode. Including patients with only a partial episode window could distort the measure by artificially reducing the actual count of patients with PACs.

(b) Medical Reasons:
Patients with outlier costs (less than 1st percentile value or greater than 99th percentile) were considered to be different from the general pool, and excluded from both the numerator and the denominator. This is another way to ensure that episodes are complete (because incomplete episodes may have very low costs), and do not bring in random noise into the analysis due to inappropriate codes or services (high outliers).

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1./S13 What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model** with 170 potential risk factors and episode specific subtypes
☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.1.2. If an outcome or resource use 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.

2b4.2/S14. Identify the statistical risk model variables (Name the statistical method – e.g., logistic regression and list all the risk factor variables.
A number of patient-related "risk factors" or covariates are included in the models:

Patient demographics: age, gender, and an indicator of whether a member has enrolled within the previous 6 months. This latter risk factor is intended to account for the patient's lack of claims history, which limits the number of potential comorbidities that can be identified.

Comorbidities: These are conditions or events that occurred prior to the start of the episode that can have a potential impact on the patient's risk of having a PAC. The risk factors are 170 disease indicators (0/1) identified through the presence of ICD diagnosis codes on individual medical claims and collected from the historical claims data before the start of an episode. These are universally applied across all episodes. Please see the tab labeled "All Risk Factors I-9" and "All Risk Factors I-10" for a list of risk factors and their corresponding codes in the enclosed workbook called NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Episode Subtypes or Severity Markers: These are markers that distinguish an episode as being more severe than another. They indicate either specific patient comorbidities that are known to make the procedure or condition more difficult to treat (e.g., obesity) or severity of the illness itself (e.g., cardiomyopathy). Subtypes are specific to each unique episode and are included in the models only if they are present at the start of the episode. Please see the tab labeled "Subtypes I-9" and "Subtypes I-10" for a list of subtypes and their corresponding codes in the enclosed workbook called NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls. This list was selected based on input from clinical experts in clinical working groups.

Candidate comorbidities and subtypes were included in the models as covariates if they were present in at least 10 episodes to prevent unstable coefficients.

2b4.2.1/S15. Detailed risk model specifications including coefficients, equations, codes with descriptors, definitions(may be attached in an Excel or cvs file)

All Risk Factors with their coefficients are detailed in the enclosed workbook called NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls – Please reference the tabs titled Risk Factor Prevalence and Risk Model.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Risk factors are comorbidity indicators collected from historical claims before the start of an episode. These are universally applied across all episodes. This list was selected based on input from clinical experts in clinical working groups. In addition, the Clinical Working Groups identified episode specific severity markers that were called episode subtypes and they help distinguish an episode as being more severe than another.

All risk factors and subtypes must be present prior to, or at the start of the episode and are identified using diagnosis codes in the patient's historical claims.

To be included in the risk adjustment models, any risk factor or subtype must be present in at least 10 episodes. Beyond this no further model building was conducted to add or remove risk factors or subtypes from the model after it was initially run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates cannot overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

As explained above, no formal analysis was conducted to select risk factors. In fact, all potential risk factors and subtypes with a count of at least 10 episodes were retained to serve as predictors. The goal was to achieve a more complete explanatory model rather than achieve parsimony.

Please reference the tabs titled Risk Model in the NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls workbook to see the list of risk factors that met the selection criteria.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Not Applicable since our analysis did include SDS variables

2b4.5. 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*)

Model Development Approach

We used logistic regression to model the probability of at least one PAC occurring during the episode. The model included all covariates that were identified through the process above. No further model building was conducted after the initial model was run. This reflects a desire to explain as much variation in the probability of having a PAC as possible, but does not make it a priority that all covariates be individually significant or even uncorrelated with each other. Accordingly, the model uses a very large group of covariates. This modeling approach allows for fewer potentially artificial constraints around the definitions of what constitutes severity, and lets the model determine for itself which of the factors are more significant. Non-significant covariates can not overly influence the predicted outcomes, nor is much harm realized, if a group of correlated covariates work together to explain variation rather than having the variation explained by a single best factor.

For a more complete description of the risk adjustment approach, please see the document entitled, "PACs and Severity Adjustment Fact Sheet" that accompanies this submission.

Approach to Model Testing and Validation

To determine the validity and performance of the model, we used the split sample method to divide the patient sample randomly into: 1) the model building data set (80% of the sample) and 2) the test data set (20% of sample. The model was built using logistic regression on the first data set and then the coefficients from the development model were tested in the second dataset. Area under the curve (AUC) and the c-statistic were used to compare the predictive ability of the model in each of the data sets. Hosmer-Lemeshow Goodness-of-Fit tests and comparisons of observed to expected probabilities across risk deciles were further examined to assess the model's overall predictive accuracy.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

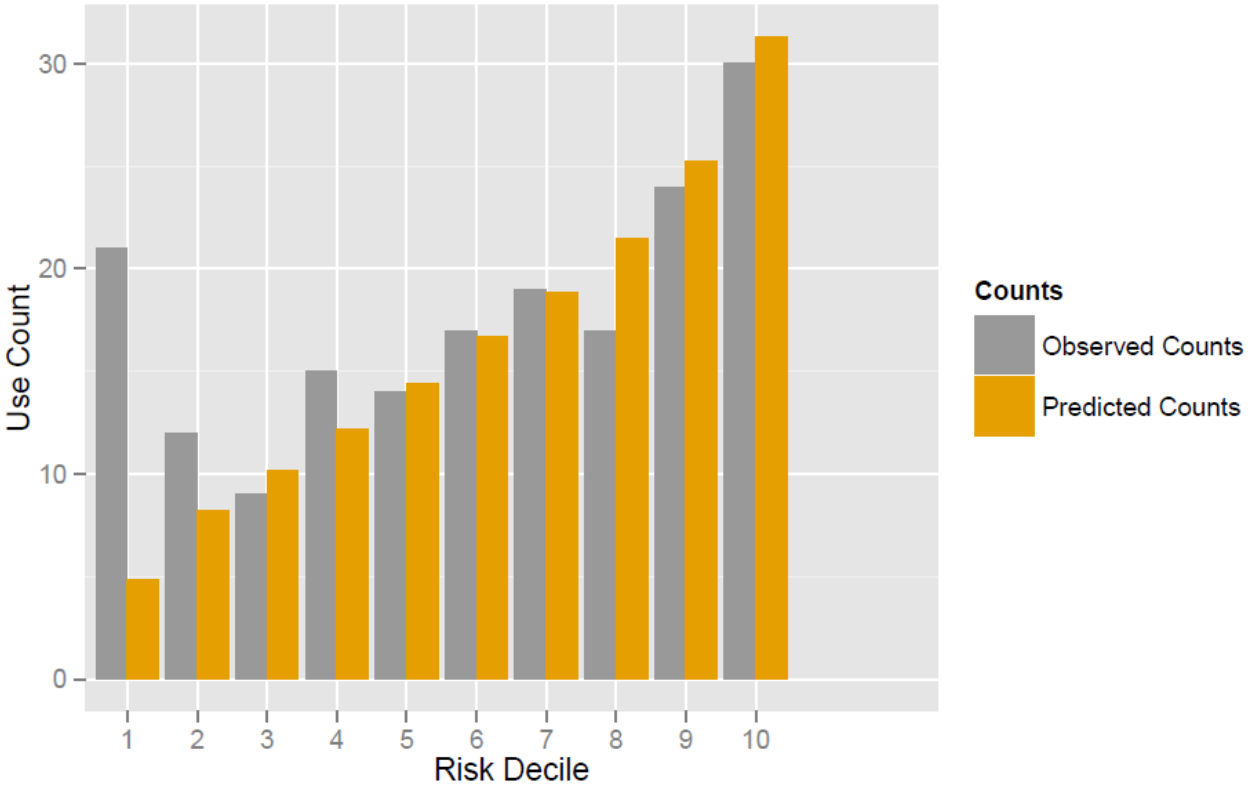
Sample	Accuracy (%)*	AUC
Test	66.8%	0.740
Validation	60.2%	0.624

*Episodes with predicted probabilities <50% were classified as having a predicted 0 (not having a PAC). Episodes with predicted probabilities >50% were classified as having a predicted 1 (having a PAC)

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Sample	Chi Square	Degrees of Freedom	p-value
Test	6.1	8	0.6338
Validation	68.7	8	<0.0001

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:



2b4.9. Results of Risk Stratification Analysis:
Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

The C statistic is a measure of the extent to which a statistical model is able to discriminate between a patient with and without an outcome. The c-statistic ranges from 0.5 to 1.0. A c-statistic of 0.50 indicates the model is no better than random prediction, implying that the patient risk factors do not predict variation in the outcome; conversely, a c-statistic of 1.0 indicates perfect prediction, implying patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play little role in patients' outcomes. Models with c-statistic values of at least 0.7 are considered good and those above 0.8 are considered strong [1]. The purpose of the model is to adjust for patient-related factors. The remaining unexplained differences in PAC rates are due to factors that could be controlled by all providers that are managing or co-managing the patient, during the entire episode time window.

The results above indicate that the C-statistic for the risk model on the testing sample (0.740) is above the level at which the model is considered to have good discriminatory power. And while the c-statistic for the testing sample was somewhat low, the model still predicted the outcome correctly 60% of the time, better than would be expected if the outcome were chosen at random (i.e., 50%). Moreover, the H-L test was not significant for the testing sample, meaning that the model was a good fit for the data. Finally, the decile plot shows that, with the exception of the first decile, the model predicts PACs somewhat similarly to observed PACs.

Overall, the results demonstrate that the model has sufficient predictive power.

Reference:
[1] Hosmer DW, Lemeshow S. *Applied Logistic Regression (2nd Edition)*. New York, NY: John Wiley & Sons; 2000.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

NA

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To directly compare PAC rates across providers or facilities while also appropriately accounting for differences in patient severity, we calculated a risk-standardized PAC rate for each provider. This method is similar to calculations used by others for reporting outcomes measures [1]. For each provider or facility, the ratio of observed attributed episodes with PACs to the expected number of attributed episodes with PACs given the patient's risk factor and estimated from the risk-adjustment model was calculated. This number yielded whether the provider or facility had more PACs than expected (ratio>1), as expected (ratio=1), or less than expected (ratio<1). We then multiplied this ratio by the overall expected PAC rate across all providers or facilities to obtain the risk-standardized PAC rate for the provider or facility. This measure represents what a facility's PAC rate would be if its patient population was reflective of the overall population.

Because facilities with small volumes may provide unreliable estimates, we excluded any with fewer than 10 attributed episodes prior to the calculations. Comparison of risk-adjusted PAC rates gives a measure of the provider's relative performance. Our analysis compared risk-standardized PAC rates across facilities. We analyzed various descriptive statistics including the range in PAC rates, medians, interquartile range, etc.

Reference:

[1] See, for example: NQF#1550: Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and / or total knee arthroplasty (TKA). Online version: http://bit.ly/1BWQTRt

2b5.2. 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*)

PAC Rates	Minimum # Episodes Per Provider	
	>=10	>=25
# Providers	22	16

Unadjusted		
Median (IQR)	47% (40%, 55%)	46% (35%, 55%)
Range	20% - 64%	26% - 64%
Adjusted (RSPR)*		
Median (IQR)	46% (37%, 54%)	46% (35%, 55%)
Range	21% - 63%	25% - 63%

***RSPR = Risk Standardized PAC Rate**

Please refer to the NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls workbook under the "ProviderAttribution Reliability" tab to see specific results for each facility.

2b5.3. 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?*)

While there was some variation among facilities in the adjusted rates, these should be considered in the context of the reliability analysis above showing that measure may not sufficiently differentiate performance differences between facilities.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved in one dataset would apply to another.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
***If only one set of specifications, this section can be skipped*.**

Note: *This criterion is directed 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 SDS 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.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not performed

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

If patient related data is missing, the entire patient is excluded from the numerator as well as the denominator.

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Furthermore, the measure is constructed so that the occurrence of any number of PACs during a defined episode would only count as one occurrence.

According to our measure definition, in constructing the measure it is possible for a provider to have only one or some types of PACs and not others. Alternatively, the provider may have all PAC types occur for their patients. The measure only considers whether any PAC occurred regardless of the type, and all PAC types are weighted equally, therefore we believe, there is no potential for the absence of specific PAC types to bias performance scores for individual providers.

For these reasons, no formal analyses were done on missing data.

2b7.2. 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*)
Not applicable

2b7.3. 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*)

Not applicable

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH
Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

The PAC measures, as we define them, look at many "care defects" comprehensively. They are composed of several cross-cutting measures and together they paint a global picture of the provider's overall performance.

PACs may occur any time during the episode time window. PACs are counted as a dichotomous (yes/no) outcome. If a patient had one or more PACs, they get counted as a "yes" or a 1. The enclosed workbook entitled NQF_PCMDFR_all_codes_risk_adjustment_06.30.15.xls provides outputs from empirical analysis. The tab labeled "PAC overview" demonstrates percentage of episodes that had at least one PAC, and provides the breakdown of PACs: 1) by the type of PAC whether directly related to index condition or due to patient safety failures; 2) the setting of the PAC, whether seen in the in-patient setting, out-patient facility or during professional visits; and 3) preventable hospitalizations.

The "PAC Drill Down Graph" provides further detail on each component of the PAC and their frequency. As can be seen by the individual counts and the graph, while each individual PAC may have such small occurrences that no meaningful comparisons in provider performances could be made; together, they add value to provide a comprehensive picture that result in meaningful numbers. The aggregation of PACs to a comprehensive, composite measure, in itself provides the parsimony that is so desirable.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

All PACs, as clinically defined by the subject matter experts were used with equal weighting. Since the emphasis of the PAC measure is to identify the occurrence of PACs in any setting, a simple and straightforward approach was adopted.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc*.; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

No formal analysis was performed.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

Since our premise is that all PACs are potentially avoidable, we adopted the approach to count all PACs and give them equal weights. The overall composite score results in the quality construct that could be measured and interpreted.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

Within our measure constructs, presence of potentially avoidable complications are identified from administrative claims data. Additionally, if a patient had one or more PACs, it is simply counted as a 1, i.e., flagged as having a PAC. The measure only considers whether any PAC occurred regardless of the type, or the site, and all PAC types are weighted equally. Therefore, no formal analysis of individual components was performed.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

We chose not to weight the components of the measure.

Considerations were given to the fact that preventable hospitalizations may be given more weight, than PACs identified in a doctor's office. Similarly PACs in an in-patient setting may have more serious implications on a patient's ultimate outcome, than PACs occurring in an outpatient setting. Additionally, preventable hospitalizations as well as index hospitalizations, each with longer lengths of stay, may have serious PACs. But how do we weigh these effects? An alternative model was considered, where cost could be considered as a surrogate for the weights. Higher cost PACs could imply more serious PACs. However, differences in costs could be driven by many issues other than the PAC itself, such as unit price of the service, method of reimbursements, contracting arrangements etc.

Furthermore, in-patient facility billing does not allow for the distinction of PAC related costs from other costs within the stay. We would fail to capture PAC related costs within the stay and potentially underweight those. As a result, the decision was made to avoid weighting and keep the measure as a straightforward count.

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i*.e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

Measuring all providers with the same yardstick will provide consistent results and reasonable comparisons over time. If the goal is to reduce PACs, then the PAC measure as was constructed with the help of various experts in the field would provide reasonable comparisons. A word of caution however pertains to the sample size of the provider panel before making any reasonable conclusions.

Minimum sample size requirements for PAC measures are a function of the reliability testing of the measures on every dataset on which the measures are applied. Our research suggests that minimum sample sizes to achieve high degrees of reliability in the measures are a function of the dataset analyzed, and as such may vary from dataset to dataset. One should not infer that a minimum sample size achieved for high reliability in one dataset would apply to another.

2d3. Empirical analysis demonstrating that the approach for handling missing data minimizes bias (*i.e., achieves scores that are an accurate reflection of quality*)**.**

Note: *Applies to the overall composite measure; the focus is on missing data rather than exclusions, which are considered in 2b3.*

Please refer to section 2b7

2d3.1. What is the overall frequency of missing data and the distribution of missing data across providers?

2d3.2. Describe the method used to compare approaches for handling missing data (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2d3.3. What were the statistical results obtained from the analysis of missing data? (e.g., *results of sensitivity analysis of effect of various rules for missing data; if no empirical analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

2d3.4. What is your interpretation of the results in terms of demonstrating that the approach used for missing data minimizes bias? (i.*e., what do the results mean in terms of supporting the selected approach for missing data; if no empirical analysis, provide rationale for the selected approach for missing data*)

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

As part of our general implementation of these measures and related analyses, we have worked through dozens of different and sometimes very large datasets. From Medicare to Medicaid to regional and national commercial carriers, as well as individual employers, the principal lesson learned is the heterogeneity of the data sets and the significant variability in fill rate of critical data elements. As a result, we have created highly specific recommendations for which data elements are required to ensure measure validity, the accuracy of those data elements, and their completeness in the dataset. When claims datasets are organized in the way we specify in the measure analysis, and contain the coding information required, the analysis of the measure and its results are highly reliable.

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 calculations of rates of potentially avoidable complications can be replicated by anyone that uses the measure specifications along with the metadata file that is available for free on our web site at http://www.hci3.org/ecre/xml-agreement.html.

We also plan on providing a limited automated analysis, at no cost, on our website.

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.

Planned	Current Use (for current use provide URL)
Public Reporting Regulatory and Accreditation Programs Professional Certification or Recognition Program	Public Health/Disease Surveillance Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/ Blue Cross Blue Shield of New Jersey http://www.horizonblue.com/ Pennsylvania Employee Benefits Trust Fund https://www.pebtf.org/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Blue Cross Blue Shield of North Carolina https://www.bcbsnc.com/assets/providers/public/pdfs/specialty_methodology.pdf

4a.1. For each CURRENT use, checked above, provide:

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

Measures associated to potentially avoidable complications (PACs) are in use today with some private sector payers and gaining further acceptance among a wide variety of organizations across the health system (public and private payers, clinicians, consultants, all-payer claims database stewards, etc.) [1-8]. They are being used in various capacities in different pilot site implementations. To name a few:

•BCBSA (Blue Cross Blue Shield Association) – uses them for their Centers of Excellence (COE) programs: Blue Distinction
•BCBSNC (Blue Cross Blue Shield of North Carolina) – is using them for tiering providers

In addition, the PAC measures are incorporated by the following organizations in their bundled payment programs:

•BCBSSC – for CABG and PCI programs
•Horizon BCBSNJ– for CHF and CABG programs
•BCBSNC
•PEBTF in PA

http://www.ajmc.com/interviews/Lili-Brillstein-on-How-Bundled-Payments-Are-Tranforming-Healthcare

In these programs they look at PACs related to the measure for process improvement activities and for practice re-engineering.

We have created reports for rates of PACs for the following organizations:
-Vermont Payment Reform
-Maine Health Management Coalition
-WellPoint / Anthem CT
-NY State Medicaid
-CT Medicaid
-CO All-payer Claims Database, Center for Improving Value in Health Care

There are several companies that are leveraging these measures to create analytics and software for customers – these include HealthQx, Aver Informatics, McKesson, and TriZetto.

Below are some references that highlight our work with Potentially Avoidable Complications (PACs).

1. Hibbard JH, Greene J, Sofaer S, Firminger K, and Hirsh J. Experiment shows that a well-designed report on costs and quality can help consumers choose high value health care. Health Affairs, 31, no.3 (2012):560-568 (doi: 10.1377/hlthaff.2011.1168)
2. Rastogi A, de Brantes F, Costley J, and Tompkins C. HCI3 Improving Incentives Issue Brief – Analysis of Medicare and Commercial Insurer-Paid Total Knee Replacement Reveals Opportunity for Cost Reduction. Available from: http://www.hci3.org/content/hci3-improving-incentives-issue-brief-analysis-medicare-and-commercial-insurer-paid-total-kn, Accessed Jun 1 2015.
3.de Brantes F, Rastogi A, and Sorensen CM. Episode of Care Analysis Reveals Sources of Variation in Costs. Am J Manag Care. 2011; 17(10): e383-e392.
4. de Brantes F, Rastogi A, and Painter M. Reducing Potentially Avoidable Complications in Patients with Chronic Diseases: The Prometheus Payment Approach. Health Services Research 2010: 45(6), Part II: 1854-1871.
5.Pierre L. Yong and LeighAnne Olsen. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary; Roundtable on Evidence-Based Medicine; Institute of Medicine. 2010. ISBN: 0-309-14434-5, http://www.nap.edu/catalog/12750.html, accessed June 14, 2015.
6.Pham HH, Ginsburg PB, Lake TK, and Maxfield MM. Episode-based Payments: Charting a course for Health care Payment Reform. National Institute for Health Care Reform. Policy Analysis, No.1. Jan 2010. Available from: http://www.nihcr.org/Episode_Based_Payments.html. Accessed Jun 1 2015.
7.François de Brantes, M.S., M.B.A., Meredith B. Rosenthal, Ph.D., and Michael Painter, J.D., M.D. Building a Bridge from Fragmentation to Accountability —The Prometheus Payment Model. NEJM 2009; 361:1033 (Perspective)
8.de Brantes F, D'Andrea G, Rosenthal MB. Should health care come with a warranty? Health Aff (Millwood) 2009; 28:w678-w687.

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

4a.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.*)
Measures associated with PACs are currently in use as described in the prior section. In addition, we are working with several not-for-profit and for-profit organizations to provide them with the algorithms needed to calculate rates of potentially avoidable complications. Some of these organizations include:

Fair Health – based in NY and whose mission is to increase transparency of provider cost and quality,

CastLight – based in CA and serving large employers. We currently provide CastLight with Bridges To Excellence recognitions and will work with them to augment provider transparency by using PAC measures,

MA APCD (Massachusetts All Payers Claims Database) Council – we currently have an agreement in place with the MA APCD Council to produce PAC measures on hospitals and physicians and report back to the council with tests of reliability and validity of the measures. The purpose is to authorize the publication of these measures,
Maryland Health Care Cost Commission – we have a two year agreement to produce measures of cost and quality for public

dissemination.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

We do not have any public information to share about the improvements in rates of potentially avoidable complications, as the implementation of these measures is too recent to provide valid comparisons. Further, some of the definitions of PACs have changed since the measures were initially endorsed, making comparisons even more difficult and unreliable.

Nevertheless, the variation in performance scores presented in Section 1b.2 indicates that there are differences between providers in their risk-adjusted PAC rates (higher scores equal worse performance). This suggests that real opportunities exist to identify lower performing providers and reduce the overall occurrence of PACs.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Performance results provide summary PACs rates by provider, which can be used by payers and providers in a number of ways to improve the quality of care.

From the payer perspective, payers can use this information to 1) create a high-value provider networks, 2) work with high-value providers to share best practices, 3) incentivize low-value providers to improve, 4) modify their insurance design to activate consumers to select the right care from the right providers at the right time.

From the provider perspective, providers can 1) view services and activity for their patients longitudinally across the entire care continuum, such as frequency of readmissions and ED visits and drill down on patients with high PAC rates, 2) review actionable drill down reports to identify the most frequent PACs across all patients to create care pathways and process improvement plans to impact the most frequent PACs.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended consequences were reported, but there is the potential for:
1. Under-coding of PACs in the claim stream resulting in under-reporting the actual rate and/or providers gaming the measures
2. Payers calculating the measures even with inadequate sample sizes and using the results to penalize providers

The measure is designed for transparency efforts and to spur quality improvement. Detailed PAC reports can help providers identify areas of quality improvement. Even detailed reports of small samples of patients can be helpful for quality improvement purposes, but not for public reporting. To mitigate the potential for invalid provider comparisons, we specify in this submission the minimum sample size needed to ensure the reliability of a provider's score. Ultimately, there isn't any good way to prevent provider gaming of the measure by under-coding claims, however, under the current DRG payment methodology, many providers would be penalized by under-coding PACs since these codes often result in the assignment of more complicated DRGs.

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)

0141 : Patient Fall Rate
0202 : Falls with injury
0337 : Pressure Ulcer Rate (PDI 2)
0450 : Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI 12)
0705 : Proportion of Patients Hospitalized with Stroke that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)
0709 : Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year.
1789 : Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

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

-0531 Patient Safety for Selected Indicators (Composite Measure, endorsed) (AHRQ)
-CMS defined hospital acquired conditions (HACs) are a subset of our PACs. We have pain-stakingly matched the definitions to provide as much consistency as possible. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html

5a. Harmonization

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 completely harmonized?

No

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

Several of the measures listed in the prior section are, in fact, fully harmonized with the submitted measures. In particular, 0705, 0708, 0709, 0531, 0450, 2503, 0337, 0141, 0202. However, there are some that are not, in particular the 30-day all-cause readmission measures. While the submitted PAC measures include readmissions that occur within 30 days of discharge, the readmissions, by definition, are related to the index hospitalization and not any hospitalization. While 30-day all-cause readmissions might make sense in a Medicare population, it is not self-evident that they do for commercial or Medicaid populations. However, that said, our data suggest that there are, in fact, very few readmissions within 30 days post discharge that aren't relevant to the index hospitalization. It is worth noting that there is some mounting controversy about the 30 day all cause readmission measures and some data suggest that these measures might have simply pushed out certain readmissions to 31 or more days post discharge. Irrespective of these points, PACs include readmissions and are designed to enable accountability at the locus of provider control as well as some shared accountability between settings, centered around a patient, and for a specific medical episode of care. In that sense, they are consistent with the all-cause 30-day readmission rates, but represent a subset of those admissions. As such, the PAC measures, as submitted, don't create added burden of reporting because the readmissions reported are simply a part of the broader 30-day all-cause readmission measures already endorsed by NQF. Because PAC measures are comprehensive, they include patient safety events that can occur during the stay, as well as adverse events, including readmissions, that can occur post-discharge. As a result, they provide facilities and physicians with an overall measure of avoidable complications for a specific medical episode. The data collection for all of the BTE measures is automated by a software package and is fully harmonized with all other PAC measures. A single download automates creation of all reports related to each of the PAC measures.

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.)
The PAC measure is a comprehensive measure representing "all-cause harms". It looks at all potentially avoidable complications in patients hospitalized with AMI during the stay or for 30-days post-discharge. It looks at readmissions, emergency room visits, adverse events due to errors of omission or commission. It looks at complications that are due to patient safety failures, and also those directly related to the index condition. These are a cause of significant waste and quality concerns for patients with an AMI episode. As such, the measure can provide clinicians with an overall and comprehensive view, in one measure, of all potentially avoidable complications for a patient and drive quality improvement efforts.

For clinicians and facilities increasingly engaged in value-based payment efforts and/or driving quality improvement for population health, the value of a PAC measure over a series of related, but more discrete measures, is that one can better determine if the sources of complications primarily stem from activities within the facility or outside the facility, and the specific nature of the complications that have a higher frequency of occurrence. For providers, it's far easier to construct a quality dashboard from a parsimonious set of measures, and that's what PAC measures offer.

Further, as a comprehensive outcome measures, PACs are also useful for public transparency of quality, as substantiated by the research from Judy Hibbard and colleagues previously cited in the "testing" section of this submission. As a comprehensive outcome measure, they are easier to explain to the average consumer. From a patient's point of view, any bad outcome has an impact on their health with respect to return to work, functional limitations and need for additional support. If a provider has a high PAC rate with regards to one component PAC but not the other PACs, the impact on the patient is still adverse. In selecting providers, individual component PAC scores would mean nothing to a patient, but aggregating it to a comprehensive quality score could be a measure of "all-cause" harms and easier to interpret and act on.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PACs_and_Severity_Adjustment_Fact_Sheet_HCI3-635719855888952322.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.2 **Point of Contact:** Francois, de Brantes, Francois.debrantes@hci3.org, 203-270-2906-
Co.3 **Measure Developer if different from Measure Steward:** Health Care Incentives Improvement Institute Inc. (HCI3)
Co.4 **Point of Contact:** Amita, Rastogi, Amita.rastogi@hci3.org, 213-934-9624-

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.
From 2006 onwards, and under the auspices of various funding organizations, HCI3 has convened and managed, or helped to convene and manage, Clinical Working Groups to inform the development and refinement of the measures. For example, in 2011, 2012 and 2013, HCI3 worked collaboratively with the American Board of Medical Specialties and the American Medical Association's Physicians Consortium for Performance Improvement, under a federal contract, to convene and get input from various clinical experts on definitions of episodes of care and their sequelae, including avoidable complications.

Some of the clinical experts that have contributed to our work include:
-Dr. John Allen, American Gastroenterology Association (AGA)

-Dr. Morton Arnsdorf, Cardiologist, University of Chicago, IL
-Dr. Peter Bach, Memorial Sloan Kettering Cancer Center (MSKCC)
-Dr. Peter Basch, Primary Care, Medstar Health, DC
-Dr. Justin Beckelman, Radiation Oncology, University of Pennsylvania, PA
-Dr. Debra Bingham, Executive Director, California Maternal Quality Care Collaborative (CMQCC) at Stanford University, CA
-Dr. John Birkmeyer, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Linda Bosserman, Wilshire Oncology Medical Group, CA
-Dr. Matthew Brengman, American Society of Metabolic and Bariatric Surgery (ASBMS)
-Dr. Joel Brill, American Gastroenterology Association (AGA)
-Dr. George Cautilli, Cautilli Orthopedic Surgical Specialists PC, Yardley, PA
-Dr. Ashwini Davison, Internist, Johns Hopkins Hospital, MD
-Dr. James Denneny, III, American Academy of Otolaryngology – Head and Neck Surgery (AAO-HNS)
-Dr. Chris Gallagher, American Society of Metabolic and Bariatric Surgery (ASMBS)
-Dr. Robert Haralson, III, American Academy of Orthopedic Surgeons (AAOS)
-Ms. Dawn Holcombe, Executive Director, Connecticut Oncology Association, CT
-Dr. Colin Howden, American Gastroenterology Association (AGA)
-Dr. John Knightly, American Association of Neurological Surgeons (AANS)
-Dr. Larry Kosinski, American Gastroenterology Association (AGA)
-Dr. Nalini Krishnan, Obstetrics & Gynecology, MN
-Dr. Kelly Kyanko, Internist, NYU School of Medicine, NY
-Dr. Tara Lagu, Internist & Infectious Disease, Baystate Medical Center, MA
-Dr. Robert Lee, Society of Thoracic Surgeons (STS)
-Dr. Alex Little, Society of Thoracic Surgeons (STS)
-Dr. Michael London, Orthopedic Surgeon, OMNI Orthopedics, OH
-Dr. Elliott Main, Obstetrics & Gynecology, California Pacific Medical Center, CA
-Dr. Constantine Mantz, 21st Century Oncology, FL
-Dr. Joseph Messer, Cardiologist, Rush University Medical Center, IL
-Dr. David Metz, American Gastroenterology Association (AGA)
-Dr. Ronald Nahass, Infectious Disease Care, NJ
-Dr. Ajay Nehra, Urologist, Rush University Medical Center, IL
-Dr. Francis Nichols, Society of Thoracic Surgeons (STS)
-Dr. Patrick O'Connor, Primary Care, HealthPartners, MN
-Dr. Sara Perkel, National Comprehensive Cancer Network, PA
-Dr. David Peura, American Gastroenterology Association (AGA)
-Dr. John Ratliff, American Association of Neurological Surgeons (AANS)
-Dr. Steven Schutzer, Connecticut Joint Replacement Institute, CT
-Dr. Leif Solberg, Primary Care, HealthPartners, MN
-Dr. Scott Sporer, Midwest Orthopedics at Rush, Chicago IL
-Dr. Bonnie Weiner, Cardiologist, Worcester Medical Center, MA
-Dr. Jonathan Weiner, Bariatric Surgery codes, Prof of Health Policy and Management, Johns Hopkins University, MD
-Dr. Janet Wright, Cardiologist, Northstate Cardiology Consultants, CA

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** Yearly
Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

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 #: 2763
De.2. **Measure Title:** Ischemic Vascular Disease Care: All or None Outcome Measure-Optimal Control
Co.1.1. **Measure Steward:** Wisconsin Collaborative for Healthcare Quality
De.3. **Brief Description of Measure:** The percentage of patients age 18 through 75 with one of the following conditions:
1) Two diagnoses related visits with Coronary Artery Disease (CAD) or a CAD risk-equivalent condition, or
2) Acute Coronary Event consisting of an acute myocardial infarction (AMI), coronary artery bypass graft (CABG), or percutaneous coronary intervention (PCI) from a hospital visit, who had each of the following during the one year measurement year:

•Documentation in the medical record of daily Aspirin or daily other antiplatelet medication usage, unless contraindicated.
•Most recent Blood pressure controlled to a level of less than 140/90 mm Hg
•Most recent Tobacco Status is Tobacco-Free
•Documentation in the medical record of Statin Use
•All or None Outcome Measure (Optimal Control) composite of BP <140/90, Tobacco Non-User, Daily Aspirin or Other Antiplatelet and Statin Use.

Patients are classified uniquely to one of the three condition subgroups in the order of Coronary Artery Disease, Coronary Artery Disease Risk-Equivalent condition, or Acute Coronary Event.
1d.3. **Developer Rationale:** Also indicated in 1d.2 above, this method was chosen because of the benefits it provides to both the patient and the practitioner. First, this methodology more closely reflects the interests and likely desires of the patient. With the data collected in a composite score, patients can easily look and see how their provider group is performing on these criteria rather than trying to make sense of multiple scores on individual measures. Second, this method represents a systems perspective emphasizing the importance of optimal care through a patient's entire healthcare experience. Third, this method gives a more sensitive scale for improvement. Whether reported at the organization, clinic site or provider level, for those scoring high marks on individual measures, the All-or-None measure will give room for benchmarks and additional improvements to be made.

Support for All-or-None measurement referenced by Nolan T, Berwick DM. All-or-None Measurement Raises the Bar on Performance. JAMA. 2006 Mar 8;295(10):1168-70.

S.4. **Numerator Statement:** All-or-None Outcome Measure (Optimal Control) - Using the IVD denominator optimal results include:
- Most recent blood pressure measurement is less than 140/90 mm Hg

And
- Most recent tobacco status is Tobacco Free

NOTE: If there is No Documentation of Tobacco Status the patient is not compliant for this measure.
And
- Daily Aspirin or Other Antiplatelet Unless Contraindicated

And
- Statin Use

S.7. **Denominator Statement:** Patients with CAD or a CAD Risk-Equivalent Condition 18-75 years of age and alive as of the last day of the MP.
S.10. **Denominator Exclusions:** There are no denominator exclusions

De.1. **Measure Type:** Composite
S.23. **Data Source:** Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry

S.26. **Level of Analysis:** Clinician : Group/Practice
Is this an eMeasure? ☐ **Yes** ☒ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☐ **No**

Is this a MAINTENANCE measure submission? ☐ **Yes** ☒ **No, this is a NEW measure submission.**
For MAINTENANCE, state the Original Endorsement Date: n/a **Most Recent Endorsement Date:** n/a
Previous Measure Evaluation: n/a

IF this measure is included in a composite, NQF Composite#/title: n/a – 2763 is the composite measure.
1d.1. **Composite Measure Construction:** two or more individual performance measure scores combined into one score
Component Measures (if endorsed or submitted for endorsement): n/a
The *non-endorsed* component measures for this composite measure included:

1. Most recent blood pressure measurement is less than 140/90 mm Hg
2. Most recent tobacco status is Tobacco Free
3. Daily Aspirin or Other Antiplatelet Unless Contraindicated
4. Statin Use

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence.

NQF considers each of the components in this composite to be an **intermediate clinical outcome**. The evidence for intermediate clinical outcomes measures should include a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured intermediate clinical outcome leads to a desired health outcome.

- This measure is a composite of intermediate outcomes intended to assess whether patients with Ischemic Vascular Disease (IVD) are receiving optimal care.
- The evidence for intermediate clinical outcomes measures should include a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measured intermediate clinical outcome leads to a desired health outcome.
- The components of this measure are:
    - Most recent blood pressure measurement is less than 140/90 mm Hg
    - Most recent tobacco status is Tobacco Free
    - Daily Aspirin or Other Antiplatelet Unless Contraindicated
    - Statin Use
- NQF criteria indicate that each component in a composite must meet the evidence subcriterion to justify its inclusion in the composite; as per NQF guidance, evidence is presented separately for each component in this measure.
- Evidence for **Aspirin/Antiplatelet Therapy**
    - Regarding the link between this intermediate outcome and patient health outcomes, the developer notes that for patients on daily aspirin or other antiplatelet, risk of further cardiovascular complications is reduced.
    - The evidence for this component is derived from the AHA/ACC Guidelines for Preventing Heart Attack and Death in Patients with Atherosclerotic Cardiovascular Disease: 2011 Update
    - The specific guideline recommendation is as follows:
        - Goal: Aspirin 75–162 mg daily is recommended in all patients with coronary artery disease

unless contraindicated
    - This is a **Class I** recommendation (defined as a condition for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective) with **level of evidence A** (defined as a recommendation based on evidence from multiple randomized trials or meta-analyses).
    - These are, respectively, the highest recommendation class and level of evidence available under the ACC/AHA grading system.
    - This recommendation is based on multiple clinical trials or meta-analyses covering multiple populations over the timespan of 2006-2011.
    - With respect to the quality, quantity, and consistency of the evidence, the developer states that there is a growing body of evidence confirming that, in patients with atherosclerotic vascular disease, comprehensive risk factor management reduces risk as assessed by a variety of outcomes, and that aspirin/antiplatelet therapy has been shown to be effective in secondary prevention of further cardiovascular risk.
    - The developer does not explicitly address the consistency across studies in the body of evidence or the estimated magnitude of effect related to aspirin/antiplatelet therapy in IVD patients.
- Evidence for **Blood Pressure Control**
    - Regarding the link between this intermediate outcome and patient health outcomes, the developer notes that blood pressure should be kept below 140mmHg systolic and 90mmHg diastolic to prevent further cardiovascular risk.
    - The evidence for this component is derived from the AHA/ACC Guidelines for Preventing Heart Attack and Death in Patients with Atherosclerotic Cardiovascular Disease: 2011 Update
    - The specific guideline recommendation is as follows:
        - Blood Pressure Goal: <140/90 mm Hg
    - This is a **Class I** recommendation (defined as a condition for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective) with **level of evidence A** (defined as a recommendation based on evidence from multiple randomized trials or meta-analyses).
    - These are, respectively, the highest recommendation class and level of evidence available under the ACC/AHA grading system.
    - This recommendation is based on multiple clinical trials or meta-analyses covering multiple populations over the timespan of 2006-2011.
    - With respect to the quality, quantity, and consistency of the evidence, the developer states that there is a growing body of evidence confirming that in patients with atherosclerotic vascular disease, comprehensive risk factor management reduces risk as assessed by a variety of outcomes, that a BP of <140/90 has been shown to prevent further cardiovascular risk, and that BP results equal to or above 140/90 should be treated, as tolerated, with blood pressure medication to achieve goal blood pressure.
    - The developer does not explicitly address the consistency across studies in the body of evidence or the estimated magnitude of effect related to maintenance of blood pressure levels below 140/90 mm Hg.
- Evidence for **Tobacco Status**
    - As evidence for this component, the developer cites a 2008 clinical practice guideline from the U.S. Department of Health And Human Services – Public Health Service.
    - The developer discusses findings from the guideline suggesting that tobacco use accounts for more than 435,000 deaths each year in the United States, noting that smoking is a known cause of multiple serious diseases and other harmful consequences.
    - The developer also notes the costs of tobacco use to society, including smoking-attributable direct medical expenses and lost productivity.
    - In addition, the developer suggests that clinicians and health care systems often fail to treat tobacco use consistently and correctly, noting that as of 2005, only 70 percent of smokers reported having received some counseling to quit, and that among current smokers who attempted to stop for at least 1 day in the past year, only 21.7 percent used cessation medication.
    - The developer suggests that there is evidence that tobacco use interventions, if delivered in a timely and effective manner, can rapidly reduce the risk of suffering from smoking-related disease.
    - Although not included by the developer, the USPSTF has made the following grade A, high certainty

recommendation for all adults: *The USPSTF recommends that clinicians ask all adults about tobacco use and provide tobacco cessation interventions for those who use tobacco products*. This USPSTF guideline goes on to note that counseling and pharmaceutical interventions increase cessation rates that and that "The USPSTF found convincing evidence that smoking cessation decreases the risk for heart disease, stroke, and lung disease."

- Evidence for **Statin Use**
  - Regarding the link between this intermediate outcome and patient health outcomes, the developer notes that for those patients on appropriate statin medication, risk of further cardiovascular complications is reduced.
  - The evidence for this component measure is derived from the 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults.
  - The specific guideline recommendation is as follows:
    - Goal: High-intensity statin therapy should be initiated or continued as first-line therapy in women and men ≤75 years of age who have clinical ASCVD, unless contraindicated.
  - This recommendation has an NHLBI grade of A (Strong), meaning there is high certainty based on evidence that the net benefit is substantial.
  - Within the ACC/AHA grading system, this is a **Class I** recommendation (defined as a condition for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective) with **level of evidence A** (defined as a recommendation based on evidence from multiple randomized trials or meta-analyses).
  - The developer notes that the evidence supporting this guideline recommendation is derived from five randomized controlled trials (RCTs), as well as systematic reviews and meta-analyses of RCTs.
  - The developer suggests that the findings of the systematic evidence review support the use of statins to prevent both nonfatal and fatal atherosclerotic cardiovascular disease (ASCVD) events, noting that the review found a high level of evidence indicating that statins reduce total mortality in individuals with a history of prior ASCVD events, and moderate evidence that statins reduce total mortality in individuals who have no prior history of ASCVD events but are at increased ASCVD risk.
  - The developer provides a table listing the effects of high-, moderate-, and low-intensity statin therapy on LDL cholesterol levels.

Questions for the Committee:

- *Is the evidence directly applicable to the intermediate outcomes being measured?*
- *Are these intermediate outcomes proximal and closely related to desired outcomes?*

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

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

- The developer presents 2014 performance data for 121 clinics, covering a total of 42,290 patients.
- Average clinic performance on the measure was .5862 (meaning that on average, clinics achieved all four goals for approximately 59 percent of eligible patients).
- Scores ranged from a minimum of .379 to a maximum of .750, with the 10th percentile at .485 and the 90th percentile at .672.
- The developer notes that this measure is not currently reported with disparities data, but that some data related to sociodemographic factors (e.g., race/ethnicity data, payer data, gender, and age) are collected and could potentially be incorporated in the future.

Questions for the Committee:

- *Do the data provided by the developer demonstrate a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

1c. Priority

1c. High Priority (previously "High Impact") requires measures to address national health goal/priority or a demonstrated high-impact aspect of care.

- *Beginning in 2015, priority is no longer an NQF measure evaluation criterion.*

1d. Composite - Quality Construct and Rationale

1c. Composite Quality Construct and Rationale. The quality construct and rationale should be explicitly articulated and logical; a description of how the aggregation and weighting of the components is consistent with the quality construct and rationale also should be explicitly articulated and logical.

- This is an 'all-or-none' composite measure, meaning if a provider does not satisfy all four components (or 'goals') for a given patient, that patient will not be counted in the composite numerator, reducing the provider's performance score accordingly.
- The quality construct for the composite is 'optimal IVD care' – using multiple dimensions of performance, the measure assesses whether patients with Ischemic Vascular Disease (IVD) are receiving optimal care.
- The developer suggests that this methodology reflects the interests and likely desires of patients—with the data collected in a single score, patients can easily look and see how their provider group is performing on these criteria rather than trying to make sense of multiple scores on individual measures.
- The developer also notes that this method represents a systems perspective, emphasizing the importance of optimal care through a patient´s entire healthcare experience.
- Finally, the developer suggests that this method provides a more sensitive instrument for assessing provider performance.

Questions for the Committee:

- *Are the quality construct and rationale for the composite explicitly stated and logical?*
- *Is the method for* aggregation and weighting of the components explicitly stated and logical?

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1. ***Committee's Overview Comments***:
   - Very strong!

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- Very strong!

1b. ***Committee's Comments on Performance Gap:***

- The developer presents 2014 performance data for 121 clinics, covering a total of 42,290 patients.
  - Average clinic performance on the measure was .5862 (meaning that on average, clinics achieved all four goals for approximately 59 percent of eligible patients).
  - Scores ranged from a minimum of .379 to a maximum of .750, with the 10th percentile at .485 and the 90th percentile at .672.
  - The developer notes that this measure is not currently reported with disparities data, but that some data related to sociodemographic factors (e.g., race/ethnicity data, payer data, gender, and age) are collected and could potentially be incorporated in the future.

1c. ***Committee's Comments on Composite Performance Measure:***

- I am worried about the following: Most recent Tobacco Status is Tobacco-Free. Health care providers have little control over this.
- They need to modify this: Documentation in the medical record of Statin Use to Documentation in the medical record of Statin Use in the absence of contraindications.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- This measure is a composite of intermediate outcomes intended to assess whether patients with Ischemic Vascular Disease (IVD) are receiving optimal care.
- The denominator population includes patients with CAD or a CAD Risk-Equivalent Condition 18-75 years of age and alive as of the last day of the measurement period.
- Only patients whose care is currently being managed within the measured entity (i.e., physician group) are included in the denominator. The developers have provided the decision logic by which these determinations are made in their submission materials, and have included CPT and HCPS codes by which office visits are identified in their code list.
- The numerator population includes denominator-eligible patients who have met each of the four intermediate outcomes that are part of this composite.
- ICD-9, ICD-10, CPT, CPT-II, and HCPS diagnosis codes used to identify the numerator and denominator populations are specified in an Excel spreadsheet included as part of the measure submission.
- For the **statin use** and **aspirin/anti-platelet** components, the developer has provided lists of specific medications that satisfy the measure's requirements.
- The measure can be reported through an all-electronic data collection method, manual review using a random-sample method (for which the developer provides an online calculator to determine appropriate sample size), or a hybrid method using administrative claims review and manual review of records where information cannot be obtained through administrative data.
- When possible, data is collected in an all-electronic format and is all-inclusive. Where there is missing data, a patient-level validation and verification process is used.
- This measure is not risk-adjusted.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- Testing data is derived from 17 group practice members of the Wisconsin Collaborative for Healthcare Quality (WCHQ).
- 15 groups reported all electronically (Total Population) and 2 groups reported using the random sample methodology.
- Testing included data from 121 clinic sites covering 50,758 patients.
- Reliability testing was performed at the measure score level.
- The developer conducted a signal-to-noise analysis of the measure score, which tests reliability by estimating the extent to which variation in scoring is caused by real differences in performance ('signal' – represented here as an estimate of provider-to-provider variance) as opposed to measurement error ('noise' – here represented as an estimate of provider-specific variance).
- Scores of signal-to-noise reliability analyses generally range from 0.0 to 1.0, with a score of zero indicating that all variation is due to measurement error and a score of 1.0 indicating that all variation is attributable to real differences across measured entities.
- Across the 121 measured clinics, average reliability was found to be **0.7817**.
- The developer states that a reliability score of greater than 0.70 is generally accepted to be sufficient for determining performance differences between groups.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- This measure is specified to determine whether patients with IVD achieved all four of the following goals:
  - Most recent blood pressure measurement is less than 140/90 mm Hg
    - Blood pressure levels are extracted directly from the patient record. If no Blood Pressure is recorded during the Measurement Period, the patient is assumed to be "not controlled."
  - Most recent tobacco status is Tobacco Free
    - Patients are asked about their tobacco status; ICD-9, CPT, HCPCS and CPT-II Codes indicating tobacco use status may also be used.
  - Daily Aspirin or Other Antiplatelet Unless Contraindicated
    - Compliance with this goal is achieved if there is either (1) documentation of an active prescription for daily Aspirin; or (2) documentation on the patient's medication list of active daily usage of Aspirin.
  - Active use of a statin unless contraindicated
    - Compliance with this goal is achieved if there is either (1) documentation of an active prescription for a statin; or (2) documentation on the patient's medication list of active usage of a statin
- The evidence presented by the developer supports:
  - Maintenance of blood pressure below 140/90 mm Hg
  - Smoking cessation in all adults
  - Aspirin 75–162 mg daily in all patients with coronary artery disease unless contraindicated
  - High-intensity statin therapy in women and men ≤75 years of age who have clinical ASCVD, unless contraindicated
- This measure is not risk-adjusted. The developer did not consider whether there was a conceptual basis for adjusting this measure for sociodemographic (SDS) factors or clinical factors.

Question for the Committee:

- *Are the specifications consistent with the evidence?*
- *Is there a conceptual basis for risk adjustment (clinical or SDS factors) of this measure?*

2b2. Validity testing

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.

- The developer has assessed this measure for data element validity using the following process:
  - The measure numerator for each reporting entity is subject to validation once every three years, on a schedule based on random selection.
  - Results that vary greatly between reporting periods or that appear significantly higher or lower than the mean are subject to validation.
  - In their first year reporting data, measured entities are validated for all measure elements (denominator and numerator).
  - Upon the release of a new measure, all reporting entities are validated for each measure element (denominator and numerator).
- The developer also conducts mapping exercises to ensure certain data elements are meeting standards and that care is being attributed to the correct providers.

- The developer notes that measure results reported through the WCHQ data repository are calculated on behalf of the reporting entity through a standard process, minimizing the chance of misinterpretation of the measure specifications by an individual reporting entity.
- Reporting entities have access to patient level data reported to the repository and are required to randomly sample patients that didn't meet the numerator to ensure that they are not missing any data elements or sources in the data files.
- Measured entities using the alternate reporting method are required to upload patient-level data files and complete on-line denominator and numerator validation forms that describe how patients met both the denominator and numerator.
- Fields required for measure validation are provided in the measure testing attachment.
- Two entities attempting to report the measure using the alternate reporting method had issues warranting removal of performance results from the public website. The issues found were related to missing medication data, data systems being combined, and staff changes in areas that work directly with the WCHQ measures. Both entities have a goal to work on resolution of the issues and plan to report the measure when it is publicly reported again in November 2015.

Questions for the Committee:

- *Does the developer's validation process ensure sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- N/A

2b4. Risk adjustment:

- This measure is not risk-adjusted.

- ***Questions for the Committee:***
  - *Do you agree with the developer that risk adjustment is not necessary for this measure?*

2b5. Meaningful difference:

- The developer did not provide any information regarding the identification of statistically significant and meaningful differences in performance; however, the reliability testing results presented in Section 2a.2 of the testing attachment suggest that the measure's ability to reliably distinguish between measured entities is adequate.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- The developer reports that if any one of the four individual component measure is missing the result (i.e., patient does not have an active aspirin order), the patient will remain in the denominator but will fail the numerator. If there is no documentation of tobacco status, the patient is not compliant for this measure.

2d.Composite measure: construction

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

- The quality construct for this composite measure is 'optimal IVD care' – using multiple dimensions of performance, the measure assesses whether patients with Ischemic Vascular Disease (IVD) are receiving optimal care.
- Because this is an all-or-none measure, empirical analysis to demonstrate that the aggregation and weighting rules are consistent with the quality construct are not needed.

Questions for the Committee:

○ *Do the component measures fit the quality construct?*

○ *Are the objectives of parsimony and simplicity achieved while supporting the quality construct?*

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- No comments

2a2.: ***Committee's Comments on Reliability-Testing:***

- Testing data is derived from 17 group practice members of the Wisconsin Collaborative for Healthcare Quality (WCHQ).
  - 15 groups reported all electronically (Total Population) and 2 groups reported using the random sample methodology.
  - Testing included data from 121 clinic sites covering 50,758 patients.
  - Reliability testing was performed at the measure score level.
  - The developer conducted a signal-to-noise analysis of the measure score, which tests reliability by estimating the extent to which variation in scoring is caused by real differences in performance ('signal' – represented here as an estimate of provider-to-provider variance) as opposed to measurement error ('noise' – here represented as an estimate of provider-specific variance).
  - Scores of signal-to-noise reliability analyses generally range from 0.0 to 1.0, with a score of zero indicating that all variation is due to measurement error and a score of 1.0 indicating that all variation is attributable to real differences across measured entities.
  - Across the 121 measured clinics, average reliability was found to be 0.7817.
  - The developer states that a reliability score of greater than 0.70 is generally accepted to be sufficient for determining performance differences between groups.

2b1.: ***Committee's Comments on Validity-Specifications***:

- What is the rationale for excluding patients older than 75 years of age?
- There is a need for more clarity regarding CAD risk-equivalent condition.

2b2.: ***Committee's Comments on Validity-Testing:***

- It is reasonable.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Risk adjustment not applicable.

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- The measure can be reported through an all-electronic data collection method based on entire IVD Denominator, a hybrid method-based on Administrative Review Denominator and Manual Review Sample, or a random sample method-based on Sample Population. The hybrid or random sample method are used if there will be missing data elements
- Data for this measure may be derived from administrative claims (CPT, CPT-II, ICD-9, ICD-10, HCPCS) or extracted directly from electronic health records.
- The developer (Wisconsin Collaborative for Healthcare Quality) also serves as a qualified clinical data registry for group practice reporting under PQRS.
- The developer states that all data elements should be available in electronic form; however, if a practice is unable to obtain the data in an all-electronic format, sampling is allowed via either hybrid or random sampling, requiring manual

chart review on the part of reporting providers, who must then submit aggregate denominator and numerator data to WCHQ's web-based reporting tool for reporting results on WCHQ's public website.

- Members of the Wisconsin Collaborative for Healthcare Quality (WCHQ) submit global patient-level files of patient demographic, encounter, and clinical data into the WCHQ data repository through a secure, HIPAA compliant portal.
- Data are obtained via data extracts (.csv files) from each practice and then uploaded into the WCHQ Repository Based Submission (RBS) database.
- Alternatively, WCHQ members can use detailed measure specifications to program measures internally and submit aggregate denominator and numerator data to WCHQ's web-based reporting tool for reporting results on WCHQ's public website. De-identified patient level data is additionally submitted for validation purposes.
- The developer has provided a spreadsheet describing the process of data submission and creation of the data files.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use at a national level?If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Committee pre-evaluation comments

Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- Yes

Criterion 4: Usability and Use

4. Usability and 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.

- This measure is being used for quality improvement with benchmarking, as well as public reporting and other accountability applications:
  - WCHQ publicly reports performance information on group practices and clinics participating in the collaborative; this includes 17 organizations reporting at the group practice level and 121 at the clinic site level, for a total of 50,758 patients.
  - WCHQ's public website also provides a Measure Summary display where members can select a reporting organization and then look at their measure results. Measure results can be viewed by Top Performer, 95th, 90th, 75th and 50th percentiles, and by Average.
  - WCHQ is a Centers for Medicare and Medicaid (CMS) approved Qualified Clinical Data Registry (QCDR), allowing eligible providers to report to WCHQ under the Physician Quality Reporting System (PQRS) starting in 2015.
- 2014 is the first year this measure was published, so the developer does not have the ability to report progress on improvement at this time. WCHQ will report the measure again in November 2015 and collect data on improvement at that time.
- The developer does not report any unintended consequences.

Questions for the Committee*:*

- *Are performance results for this measure being used to further the goal of high-quality, efficient healthcare?*
- *Are you aware of any potential unintended consequences of this measure?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- This measure is being used for quality improvement with benchmarking, as well as public reporting and other accountability applications:
  - WCHQ publicly reports performance information on group practices and clinics participating in the collaborative; this includes 17 organizations reporting at the group practice level and 121 at the clinic site level, for a total of 50,758 patients.
  - WCHQ's public website also provides a Measure Summary display where members can select a reporting organization and then look at their measure results. Measure results can be viewed by Top Performer, 95th, 90th, 75th and 50th percentiles, and by Average.
  - WCHQ is a Centers for Medicare and Medicaid (CMS) approved Qualified Clinical Data Registry (QCDR), allowing eligible providers to report to WCHQ under the Physician Quality Reporting System (PQRS) starting in 2015.

Criterion 5: Related and Competing Measures

- List any related or competing measures based on harmonization protocol.
- Summarize any harmonization efforts, i.e., responses from the developers regarding harmonization.
- Briefly summarize next steps according to protocol

Pre-meeting public and member comments

Comment by: Ashish R. Trivedi, Pharm.D.
Organization: SPI
Comment#5116: "While Lilly is supportive of this measure, we suggest the use of dual anti-platelet therapy (treatment with aspirin and a P2Y12 inhibitor) as supported by the treatment guidelines for patients with acute coronary syndrome (ACS, including AMI) and/or those managed with revascularization [O'Gara et al 2013, Amsterdam et al, 2014, Levine et al, 2011].

References

- O'Gara PT, Kushner FG, Ascheim DD, et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation. 2013;127(4):e362-425. doi:10.1161/CIR.0b013e3182742cf6.
- Amsterdam EA, Wenger NK, Brindis RG, Casey DE, Jr., Ganiats TG, Holmes DR, Jr. et al. 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;130(25):e344-426. doi:10.1161/CIR.0000000000000134.
- Levine GN, Bates ER, Blankenship JC, et al. 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. Circulation.2011;124(23):e574-651. doi: 10.1161/CIR.0b013e31823ba622."

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Component # 1 DAILY ASPIRIN OR OTHER ANTIPLATELET MEDICATIONS THERAPY UNLESS CONTRAINDICATED

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Ischemic Vascular Disease Care: Daily Aspirin or Other Antiplatelet Medications Therapy Unless Contraindicated Component

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Ischemic Vascular Disease Care: All or None Outcome Measure-Optimal Control

Date of Submission: 6/29/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☒ Intermediate clinical outcome (*e.g., lab value*): Daily Aspirin or Antiplatelet Use Unless Contraindicated for patients with ischemic vascular disease (IVD)

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

- Determine patients with diagnosis of IVD
- Assess patients with IVD that are on a daily aspirin or other antiplatelet unless contraindicated
- For those patients on daily aspirin or other antiplatelet risk of further cardiovascular complications is reduced.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

B/P Control, Tobacco Cessation and Daily Aspirin Guideline:

Sidney C. Smith, Jr, MD; Steven N. Blair, PED; Robert O. Bonow, MD; Lawrence M. Brass, MD; Manuel D. Cerqueira, MD; Kathleen Dracup, RN, DNSc; Valentin Fuster, MD, PhD; Antonio Gotto, MD, DPhil; Scott M. Grundy, MD, PhD; Nancy Houston Miller, RN, BSN; Alice Jacobs, MD; Daniel Jones, MD; Ronald M. Krauss, MD; Lori Mosca, MD, PhD; Ira Ockene, MD; Richard C. Pasternak, MD; Thomas Pearson, MD, PhD; Marc A. Pfeffer, MD, PhD; Rodman D. Starke, MD; Kathryn A. Taubert, PhD

AHA/ACC Guidelines for Preventing Heart Attack and Death in Patients with Atherosclerotic Cardiovascular Disease: 2001 Update (November 1, 2001)-

http://circ.ahajournals.org/content/124/22/2458.full.pdf+html

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Antiplatelet agents/anticoagulants	**Class I** 1. Aspirin 75–162 mg daily is recommended in all patients with coronary artery disease unless contraindicated.[64,81,82,116] ***(Level of Evidence: A)***

Daily Aspirin Guideline: Page 3: Goal - Aspirin 75–162 mg daily is recommended in all patients with coronary artery disease unless contraindicated

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Daily Aspirin Guideline: Class I, Level of Evidence A – Recommendation that treatment or procedure is effective. Sufficient evidence from multiple randomized trials or meta-analyses.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

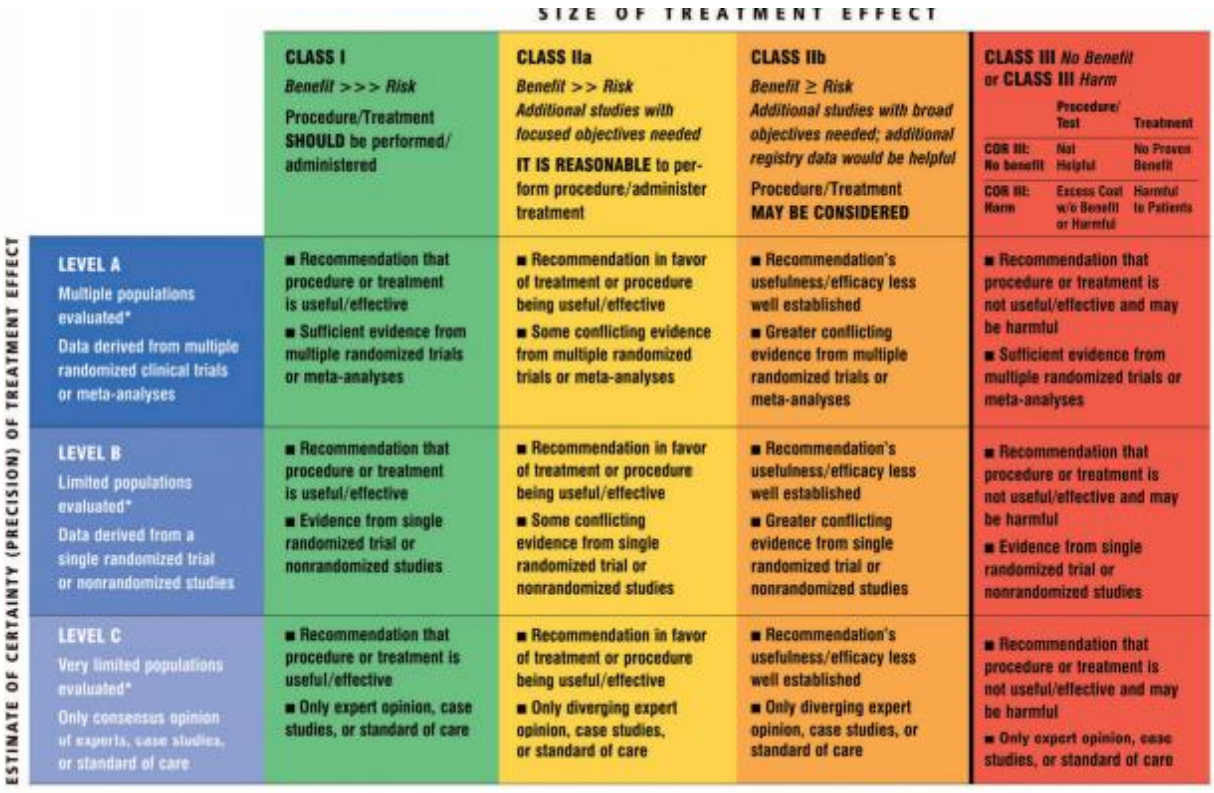
For more than 20 years, the American College of Cardiology (ACC) and the American Heart Association (AHA) have released clinical practice guidelines to provide recommendations on care of patients with cardiovascular disease. The ACC/AHA guidelines currently use a grading schema based on level of evidence and class of recommendation (available at *http://www.acc.org* and *http://www.aha.org*). The level of evidence classification combines an objective description of the existence and the types of studies supporting the recommendation and expert consensus, according to 1 of the following 3 categories:

- Level of evidence A: recommendation based on evidence from multiple randomized trials or meta-analyses
- Level of evidence B: recommendation based on evidence from a single randomized trial or nonrandomized studies
- Level of evidence C: recommendation based on expert opinion, case studies, or standards of care.

The class of recommendation designation indicates the strength of a recommendation and requires guideline writers not only to make a judgment about the relative strengths and weaknesses of the study data but also to make a value

judgment about the relative importance of the risks and benefits identified by the evidence and to synthesize conflicting findings among multiple studies. Definitions of the classes of recommendation are as follows:

- Class I: conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective
- Class II: conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment
- Class IIa: weight of evidence/opinion is in favor of usefulness/efficacy
- Class IIb: usefulness/efficacy is less well established by evidence/opinion
- Class III: conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective and in some cases may be harmful.



1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

See 1a.4.1

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

Antiplatelet agents/anticoagulants	**Class I** 1. Aspirin 75–162 mg daily is recommended in all patients with coronary artery disease unless contraindicated.[64,81,82,116] ***(Level of Evidence: A)***

Daily Aspirin Guideline: Page 3: Goal - Aspirin 75–162 mg daily is recommended in all patients with coronary artery disease unless contraindicated

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Daily aspirin or antiplatelet unless contraindicated for patients with IVD.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Class 1 – Level of Evidence A.

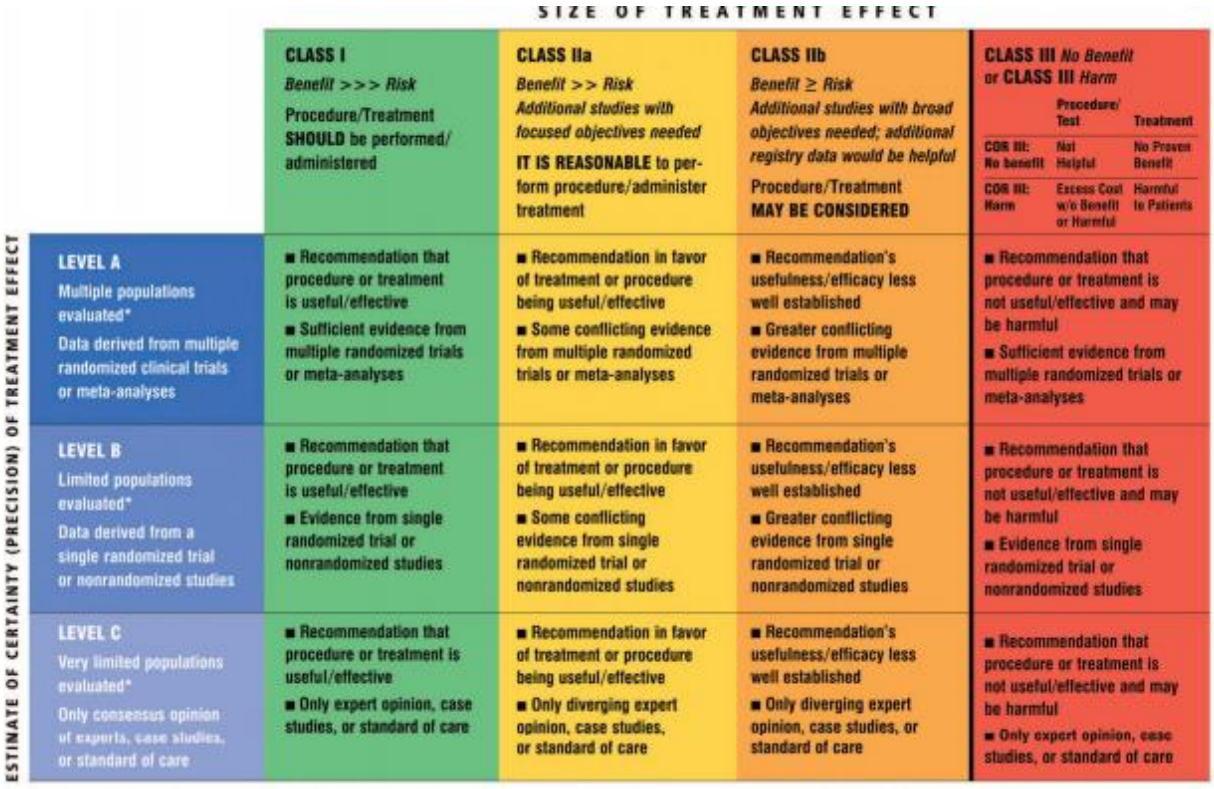
1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

For more than 20 years, the American College of Cardiology (ACC) and the American Heart Association (AHA) have released clinical practice guidelines to provide recommendations on care of patients with cardiovascular disease. The ACC/AHA guidelines currently use a grading schema based on level of evidence and class of recommendation (available at *http://www.acc.org* and *http://www.aha.org*). The level of evidence classification combines an objective description of the existence and the types of studies supporting the recommendation and expert consensus, according to 1 of the following 3 categories:

- Level of evidence A: recommendation based on evidence from multiple randomized trials or meta-analyses
- Level of evidence B: recommendation based on evidence from a single randomized trial or nonrandomized studies
- Level of evidence C: recommendation based on expert opinion, case studies, or standards of care.

The class of recommendation designation indicates the strength of a recommendation and requires guideline writers not only to make a judgment about the relative strengths and weaknesses of the study data but also to make a value judgment about the relative importance of the risks and benefits identified by the evidence and to synthesize conflicting findings among multiple studies. Definitions of the classes of recommendation are as follows:

- Class I: conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective
- Class II: conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment
- Class IIa: weight of evidence/opinion is in favor of usefulness/efficacy
- Class IIb: usefulness/efficacy is less well established by evidence/opinion
- Class III: conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective and in some cases may be harmful.



1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 2006-2011

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Multiple populations evaluated in multiple clinical trials or meta-analysis

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

The growing body of evidence confirms that in patients with atherosclerotic vascular disease, comprehensive risk factor management reduces risk as assessed by a variety of outcomes, including improved survival, reduced recurrent events, the need for revascularization procedures, and improved quality of life. It is important not only that the healthcare provider implement these recommendations in appropriate patients but also that healthcare systems support this implementation to maximize the benefit to the patient. Compelling evidence-based results from recent clinical trials and revised practice guidelines provide the impetus for this update of the 2006 recommendations with evidence-based results (see 1a.7). Classification of recommendations and level of evidence are expressed in ACCF/AHA format, as detailed in Table 2 (see table in 1a.73). Recommendations made herein are largely based on major practice guidelines from the National Institutes of Health and updated ACCF/AHA practice guidelines, as well as on results from recent clinical trials. Writing group members performed relevant supplemental literature searches on pertinent topics in this guideline. Additional searches cross-referenced these topics with the

subtopics of clinical trials, secondary prevention, atherosclerosis, and coronary/cerebral/peripheral artery disease. These searches were limited to studies, reviews, and other evidence conducted in human subjects and published in English. In addition, the writing group reviewed documents related to the subject matter previously published by the AHA, the ACCF, and the National Institutes of Health.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Aspirin/antiplatelet therapy has been shown to be effective in secondary prevention of further cardiovascular risk.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

Component # 2 BLOOD PRESSURE CONTROL

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Ischemic Vascular Disease Care: Blood Pressure Control Component

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Ischemic Vascular Disease Care: All or None Outcome Measure-Optimal Control

Date of Submission: 6/29/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

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

Outcome

☐ Health 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

☒ Intermediate clinical outcome (*e.g., lab value*): Blood Pressure Control

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

- Assess Blood Pressure at each health care encounter
- Blood Pressure result should be less than 140 mmHg systolic and less than 90mmHg diastolic
- All patients with a BP of >=130/90 should be counseled regarding the need for lifestyle modification: weight control; increased physical activity; alcohol moderation; sodium reduction; and emphasis on increased consumption of fresh fruits, vegetables, and low-fat dairy products
- If result is 140/90 or greater patient should be treated, as tolerated, with blood pressure medication to achieve goal blood pressure
- Secondary prevention of further cardiovascular risk

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

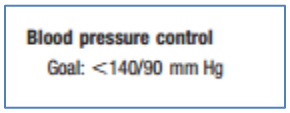
1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Sidney C. Smith, Jr, MD; Steven N. Blair, PED; Robert O. Bonow, MD; Lawrence M. Brass, MD; Manuel D. Cerqueira, MD; Kathleen Dracup, RN, DNSc; Valentin Fuster, MD, PhD; Antonio Gotto, MD, DPhil; Scott M. Grundy, MD, PhD; Nancy Houston Miller, RN, BSN; Alice Jacobs, MD; Daniel Jones, MD; Ronald M. Krauss, MD; Lori Mosca, MD, PhD; Ira Ockene, MD; Richard C. Pasternak, MD; Thomas Pearson, MD, PhD; Marc A. Pfeffer, MD, PhD; Rodman D. Starke, MD; Kathryn A. Taubert, PhD

AHA/ACC Guidelines for Preventing Heart Attack and Death in Patients with Atherosclerotic Cardiovascular Disease: 2001 Update (November 1, 2001)-

http://circ.ahajournals.org/content/124/22/2458.full.pdf+html

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.



B/P Control Guideline : Page 2: Goal - 140/90 mm Hg

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

B/P Control Guideline : Class I, Level of Evidence A – Recommendation that treatment or procedure is effective. Sufficient evidence from multiple randomized trials or meta-analyses.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

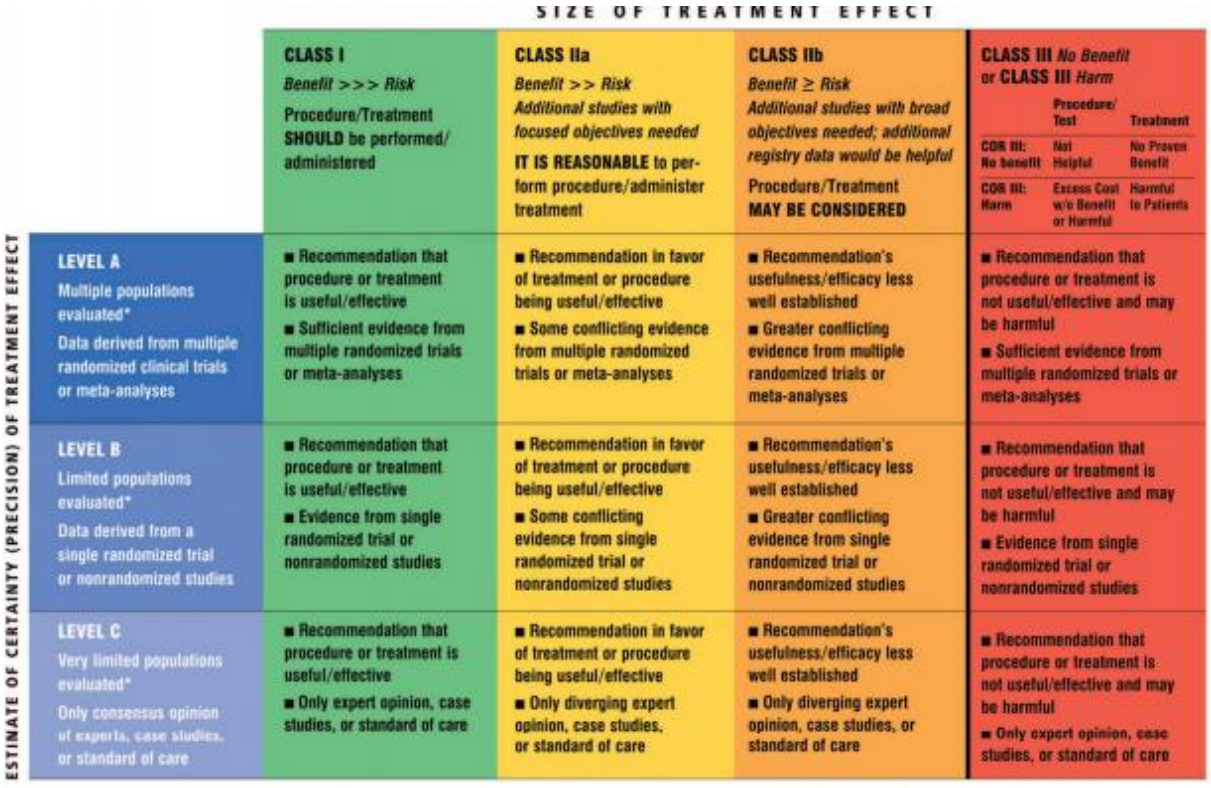
For more than 20 years, the American College of Cardiology (ACC) and the American Heart Association (AHA) have released clinical practice guidelines to provide recommendations on care of patients with cardiovascular disease. The ACC/AHA guidelines currently use a grading schema based on level of evidence and class of recommendation (available at *http://www.acc.org* and *http://www.aha.org*). The level of evidence classification combines an objective description

of the existence and the types of studies supporting the recommendation and expert consensus, according to 1 of the following 3 categories:

- Level of evidence A: recommendation based on evidence from multiple randomized trials or meta-analyses
- Level of evidence B: recommendation based on evidence from a single randomized trial or nonrandomized studies
- Level of evidence C: recommendation based on expert opinion, case studies, or standards of care.

The class of recommendation designation indicates the strength of a recommendation and requires guideline writers not only to make a judgment about the relative strengths and weaknesses of the study data but also to make a value judgment about the relative importance of the risks and benefits identified by the evidence and to synthesize conflicting findings among multiple studies. Definitions of the classes of recommendation are as follows:

- Class I: conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective
- Class II: conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment
- Class IIa: weight of evidence/opinion is in favor of usefulness/efficacy
- Class IIb: usefulness/efficacy is less well established by evidence/opinion
- Class III: conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective and in some cases may be harmful.



1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

See 1a.4.1

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

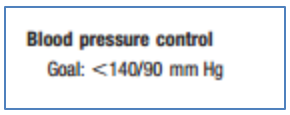
1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.



B/P Control Guideline : Page 2: Goal - 140/90 mm Hg

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Blood Pressure Control Goal less than 140 mmHg systolic and less than 90 mmHg diastolic

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

B/P Control Guideline : Class I, Level of Evidence A – Recommendation that treatment or procedure is effective. Sufficient evidence from multiple randomized trials or meta-analyses.

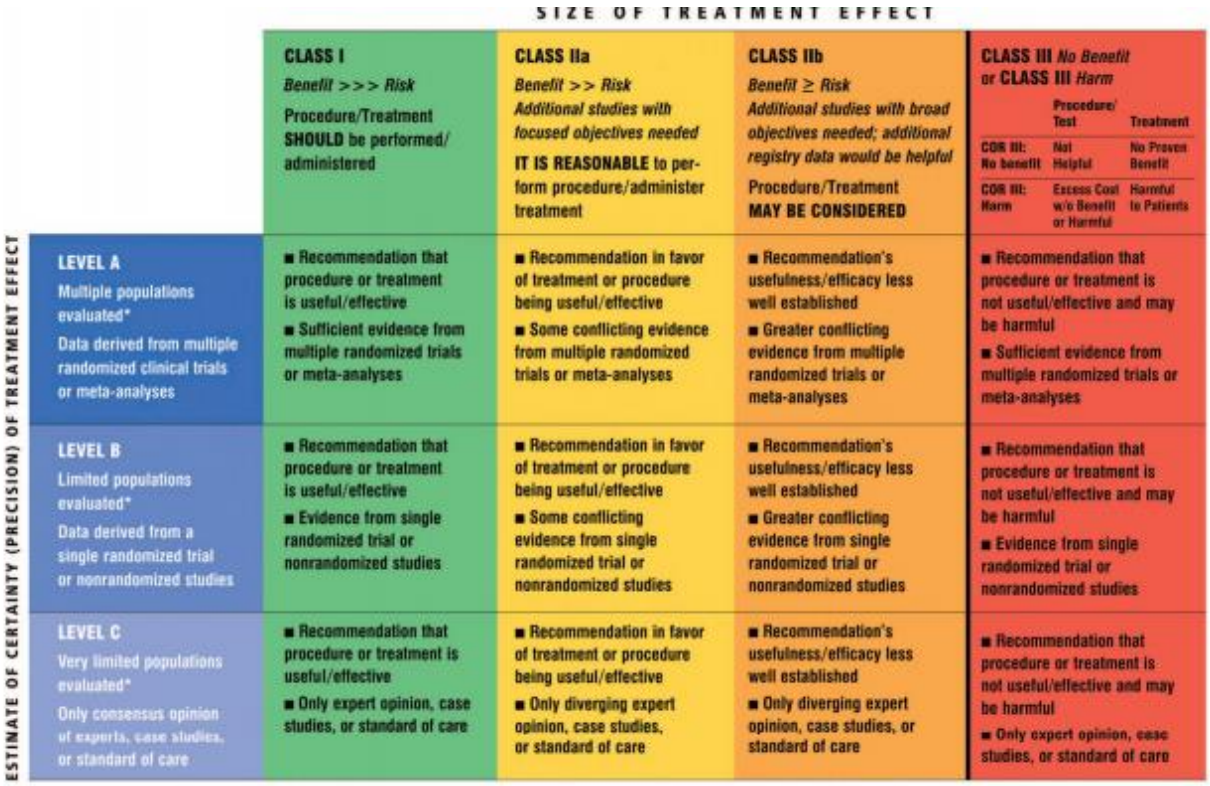
1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

For more than 20 years, the American College of Cardiology (ACC) and the American Heart Association (AHA) have released clinical practice guidelines to provide recommendations on care of patients with cardiovascular disease. The ACC/AHA guidelines currently use a grading schema based on level of evidence and class of recommendation (available at *http://www.acc.org* and *http://www.aha.org*). The level of evidence classification combines an objective description of the existence and the types of studies supporting the recommendation and expert consensus, according to 1 of the following 3 categories:

- Level of evidence A: recommendation based on evidence from multiple randomized trials or meta-analyses
- Level of evidence B: recommendation based on evidence from a single randomized trial or nonrandomized studies
- Level of evidence C: recommendation based on expert opinion, case studies, or standards of care.

The class of recommendation designation indicates the strength of a recommendation and requires guideline writers not only to make a judgment about the relative strengths and weaknesses of the study data but also to make a value judgment about the relative importance of the risks and benefits identified by the evidence and to synthesize conflicting findings among multiple studies. Definitions of the classes of recommendation are as follows:

- Class I: conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective
- Class II: conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment
- Class IIa: weight of evidence/opinion is in favor of usefulness/efficacy
- Class IIb: usefulness/efficacy is less well established by evidence/opinion
- Class III: conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective and in some cases may be harmful.



1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 2006-2011

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Multiple populations evaluated in multiple clinical trials or meta-analysis

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

The growing body of evidence confirms that in patients with atherosclerotic vascular disease, comprehensive risk factor management reduces risk as assessed by a variety of outcomes, including improved survival, reduced recurrent events, the need for revascularization procedures, and improved quality of life. It is important not only that the healthcare provider implement these recommendations in appropriate patients but also that healthcare systems support this implementation to maximize the benefit to the patient. Compelling evidence-based results from recent clinical trials and revised practice guidelines provide the impetus for this update of the 2006 recommendations with evidence-based results (see 1a.7). Classification of recommendations and level of evidence are expressed in ACCF/AHA format, as detailed in Table 2 (see table in 1a.73). Recommendations made herein are largely based on major practice guidelines from the National Institutes of Health and updated ACCF/AHA practice guidelines, as well as on results from recent clinical trials. Writing group members performed relevant supplemental literature searches on pertinent topics in this guideline. Additional searches cross-referenced these topics with the subtopics of clinical trials, secondary prevention, atherosclerosis, and coronary/cerebral/peripheral artery

disease. These searches were limited to studies, reviews, and other evidence conducted in human subjects and published in English. In addition, the writing group reviewed documents related to the subject matter previously published by the AHA, the ACCF, and the National Institutes of Health.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

In patients with IVD a BP of <140/90 has been shown to prevent further cardiovascular risk. Results equal to or above 140/90 should be treated, as tolerated, with blood pressure medication to achieve goal blood pressure.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

None

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Component # 3 TOBACCO FREE

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Ischemic Vascular Disease Care: Tobacco Free Component

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Ischemic Vascular Disease Care: All or None Outcome Measure-Optimal Control

Date of Submission: 6/29/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient

input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

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

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

Outcome

☒ Health outcome: Patient is tobacco free

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

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

- Ask patient about tobacco status at each office visit
- If patient is a tobacco-user (cigarettes, pipe, smokeless, etc.) provide cessation advice/counseling
- Provide pharmacological therapy (i.e. medication, patch) if patient is willing
- Provide formal cessation counseling
- Desired outcome is that patient will be tobacco free

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

http://www.ctri.wisc.edu/Researchers/Guideline_update/cpg_full2008.pdf

According to the 2008 Clinical Practice Guideline from the U.S. Department of Health and Human Services "Treating Tobacco Use and Dependence" Tobacco use has been cited as the chief avoidable cause of illness and death in our society and accounts for more than 435,000 deaths each year in the United States.37,38 Smoking is a known cause of multiple cancers, heart disease, stroke, complications of pregnancy, chronic obstructive pulmonary disease (COPD), and many other diseases. In addition, recent research has documented the substantial health dangers of involuntary exposure to tobacco smoke. Despite these health dangers and the public's awareness of those dangers, tobacco use remains surprisingly prevalent. Recent estimates are that about 21 percent of adult Americans smoke,3 representing approximately 45 million current adult smokers. Moreover, tobacco use remains a pediatric disease. Each day, about 4,000 youth ages 12 to 17 years smoke their first cigarette, and about 1,200 children and adolescents become daily cigarette smokers. As a result, new generations of Americans are at risk for the extraordinarily harmful consequences of tobacco use.

Tobacco use exacts a heavy cost to society as well as to individuals. Smoking-attributable health care expenditures are estimated at $96 billion per year in direct medical expenses and $97 billion in lost productivity. It has been estimated that the per pack additional cost of smoking to society is approximately $7.18 per pack, and the combined cost of each pack to society and the individual smoker and family is nearly $40.46 If all smokers covered by state Medicaid programs quit, the annual savings to Medicaid would be $9.7 billion after 5 years.

Despite the tragic consequences of tobacco use, clinicians and health care systems often fail to treat it consistently and effectively. For instance, in 1995, about the time of the release of the first clinical practice guideline, smoking status was identified in only about 65 percent of clinic visits, and smoking cessation counseling was provided in only 22 percent of smokers' clinic visits. Moreover, treatment typically was offered only to patients already suffering from tobacco-related diseases. This pattern gradually began to improve as of 2005, with up to 90 percent of smokers reporting they had been asked about smoking status and more than 70 percent reporting having received some counseling to quit. However, the failure to assess and intervene consistently with all tobacco users continues despite sub-stantial evidence that even brief interventions can be effective among many different populations of smokers. Also, the use of effective medications is low. Among current smokers who attempted to stop for at least 1 day in the past year, only 21.7 percent used cessation medication.

This Guideline concludes that tobacco use presents a rare confluence of circumstances: (1) a highly significant health threat; (2) a lack of consistent intervention by clinicians; and (3) the presence of effective interventions. This last point is buttressed by evidence that tobacco use interventions, if delivered in a timely and effective manner, can rapidly reduce the risk of suffering from smoking-related disease. Indeed, it is difficult to identify any other condition that presents such a mix of lethality, prevalence, and neglect, despite effective and readily available interventions.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

Component # 4 STATIN USE

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Ischemic Vascular Disease Care: Statin Use Component

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Ischemic Vascular Disease Care: All or None Outcome Measure-Optimal Control

Date of Submission: 6/29/2015

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

Note: The information provided in this form is intended to aid the Steering 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:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

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

Outcome

☐ Health 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

☒ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

- Determine patients with diagnosis of IVD
- Assess patients with IVD that are on a statin medication
- If patient has IVD and is not on a statin medication, prescribe an appropriate medication, if it can be tolerated
- For those patients on appropriate statin medication risk of further cardiovascular complications is reduced.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

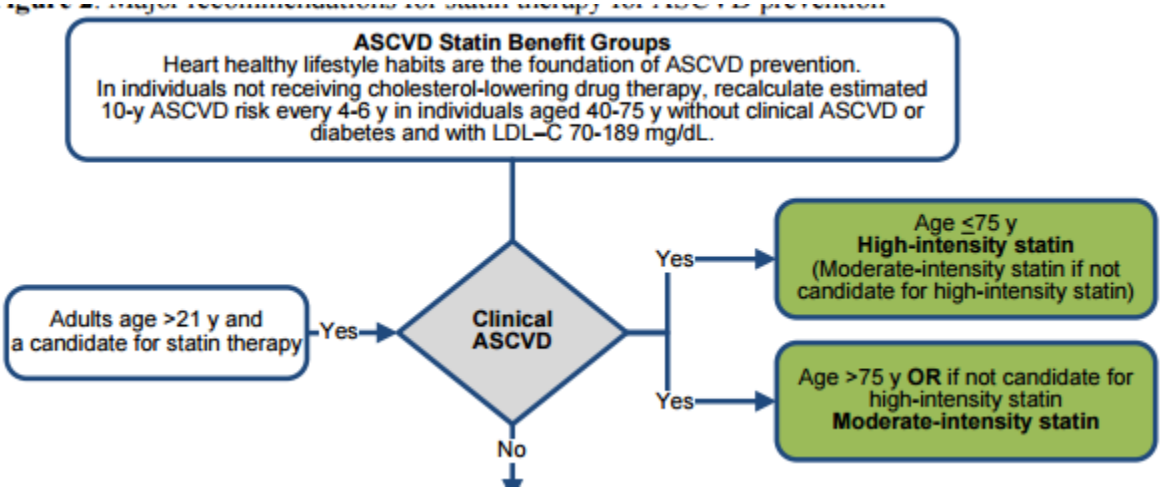
1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Blood Cholesterol Guideline:

Stone NJ, Robinson J, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero ST, Smith SC Jr, Watson K, Wilson PWF. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2013;00:000–000.

https://circ.ahajournals.org/content/early/2013/11/11/01.cir.0000437738.63853.7a.full.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.



The statin RCTs provide the most extensive evidence for the greatest magnitude of ASCVD event reduction, with the best margin of safety. Identification of 4 Statin Benefit Groups - in which the potential for an ASCVD risk reduction benefit clearly exceeds the potential for adverse effects in individuals with clinical ASCVD (only one benefit group shown here, pertinent to this measure of IVD)

AHA/ACC Blood Cholesterol Guideline: Secondary Prevention #1, page 16: Goal - High-intensity statin therapy should be initiated or continued as first-line therapy in women and men ≤75 years of age who have clinical ASCVD*, unless contraindicated.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

NHLBI Grade A (Strong) - There is high certainty based on evidence that the net benefit† is substantial.

Recommendations	NHLBI Grade	NHLBI Evidence Statements	ACC/AHA COR	ACC/AHA LOE
Secondary Prevention				
1. High-intensity statin therapy should be initiated or continued as first-line therapy in women and men ≤75 years of age who have *clinical ASCVD**, unless contraindicated.	A (Strong)	1, 6-8, 10-23, 26-28	I	A

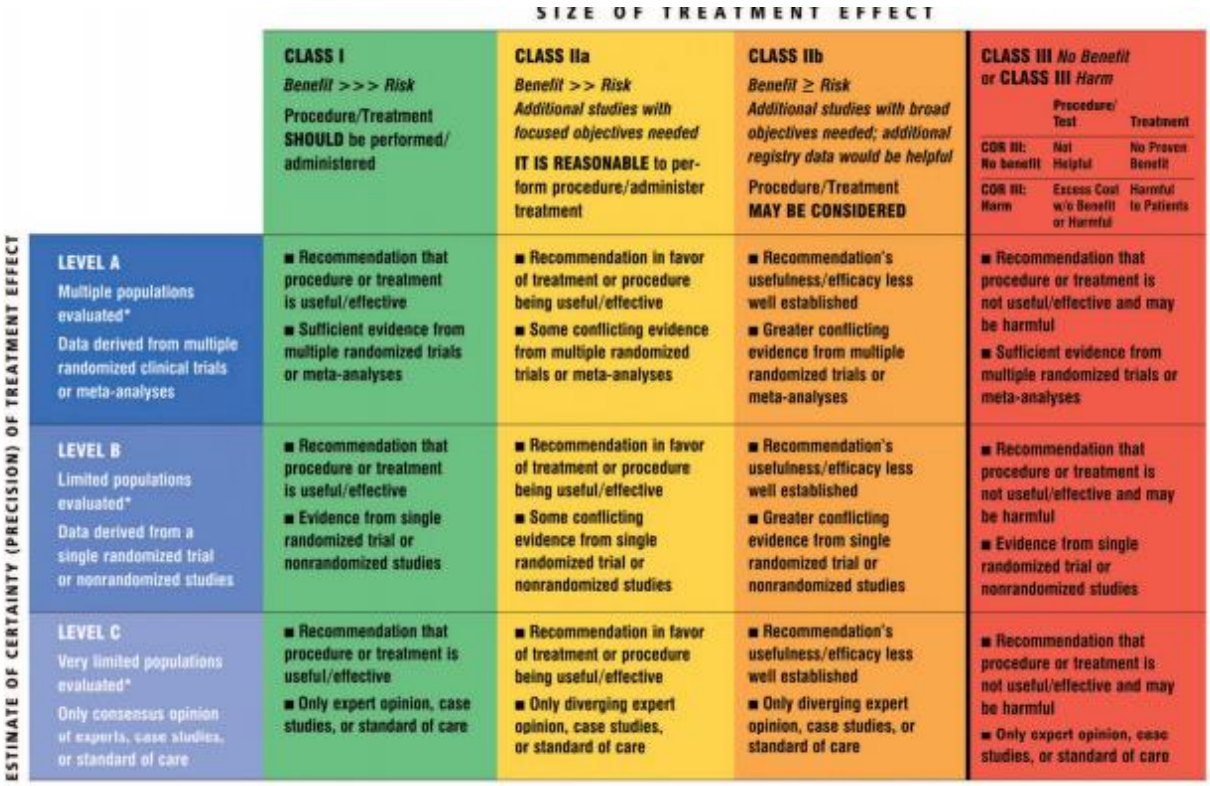
1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

See tables below:

Table 1a. NHLBI Grading the Strength of Recommendations

Grade	Strength of Recommendation*
A	**Strong recommendation** There is high certainty based on evidence that the net benefit† is substantial.
B	**Moderate recommendation** There is moderate certainty based on evidence that the net benefit is moderate to substantial, or there is high certainty that the net benefit is moderate.
C	**Weak recommendation** There is at least moderate certainty based on evidence that there is a small net benefit.
D	**Recommendation against** There is at least moderate certainty based on evidence that it has no net benefit or that risks/harms outweigh benefits.
E	**Expert opinion ("There is insufficient evidence or evidence is unclear or conflicting, but this is what the Work Group recommends.")** Net benefit is unclear. Balance of benefits and harms cannot be determined because of no evidence, insufficient evidence, unclear evidence, or conflicting evidence, but the Work Group thought it was important to provide clinical guidance and make a recommendation. Further research is recommended in this area.
N	**No recommendation for or against ("There is insufficient evidence or evidence is unclear or conflicting.")** Net benefit is unclear. Balance of benefits and harms cannot be determined because of no evidence, insufficient evidence, unclear evidence, or conflicting evidence, and the Work Group thought no recommendation should be made. Further research is recommended in this area.

*In most cases, the strength of the recommendation should be closely aligned with the quality of the evidence;



1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

See 1a.4.1

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

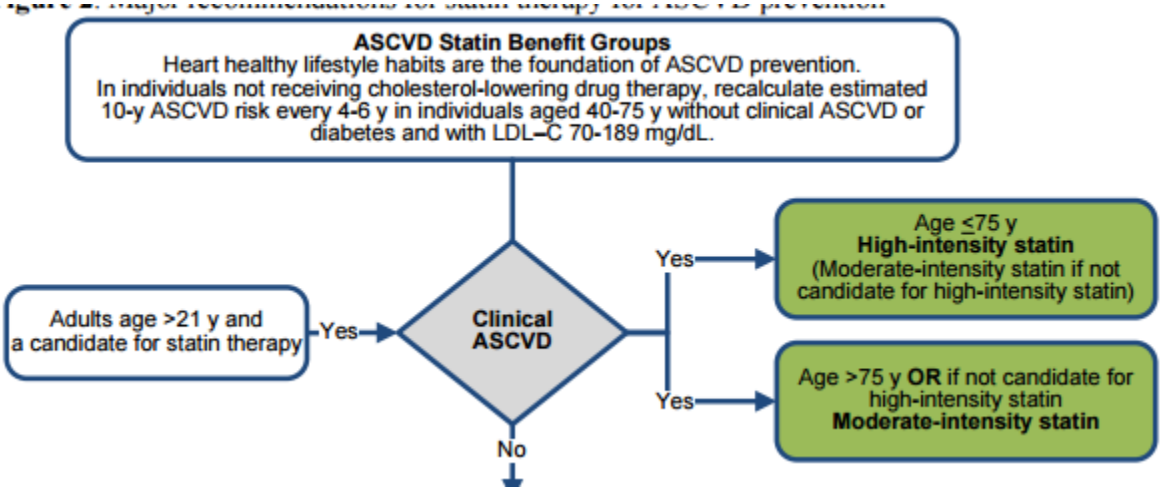
1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.



The statin RCTs provide the most extensive evidence for the greatest magnitude of ASCVD event reduction, with the best margin of safety. Identification of 4 Statin Benefit Groups - in which the potential for an ASCVD risk reduction benefit clearly exceeds the potential for adverse effects in individuals with clinical ASCVD. (only one benefit group shown here, pertinent to this measure of IVD)

AHA/ACC Blood Cholesterol Guideline: Secondary Prevention #1, page 16: Goal - High-intensity statin therapy should be initiated or continued as first-line therapy in women and men ≤75 years of age who have clinical ASCVD*, unless contraindicated.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

High-intensity statin therapy should be initiated or continued as first-line therapy in women and men ≤75 years of age who have clinical ASCVD*, unless contraindicated.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

NHLBI Grade A (Strong) - There is high certainty based on evidence that the net benefit† is substantial.

Recommendations	NHLBI Grade	NHLBI Evidence Statements	ACC/AHA COR	ACC/AHA LOE
Secondary Prevention				
1. High-intensity statin therapy should be initiated or continued as first-line therapy in women and men ≤75 years of age who have *clinical ASCVD**, unless contraindicated.	A (Strong)	1, 6-8, 10-23, 26-28	I	A

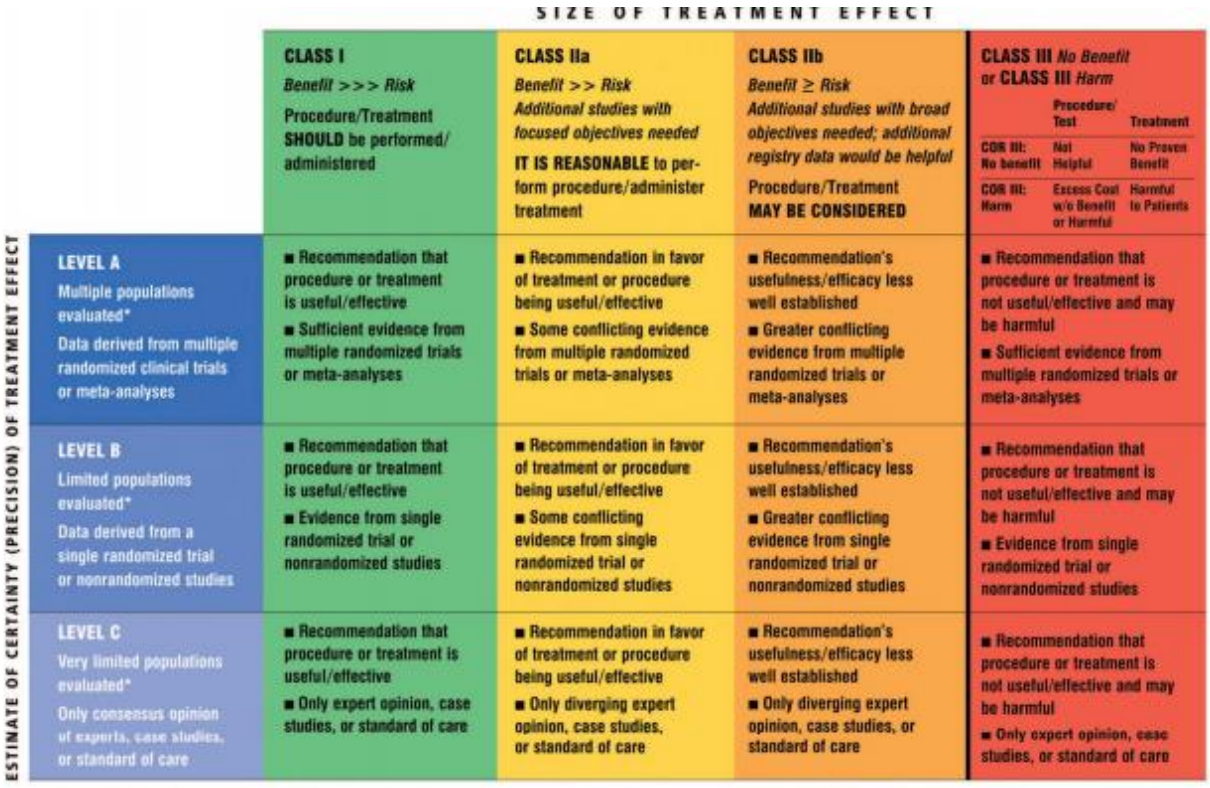
1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Table 1a. NHLBI Grading the Strength of Recommendations

Grade	Strength of Recommendation*
A	**Strong recommendation** There is high certainty based on evidence that the net benefit† is substantial.
B	**Moderate recommendation** There is moderate certainty based on evidence that the net benefit is moderate to substantial, or there is high certainty that the net benefit is moderate.
C	**Weak recommendation** There is at least moderate certainty based on evidence that there is a small net benefit.
D	**Recommendation against** There is at least moderate certainty based on evidence that it has no net benefit or that risks/harms outweigh benefits.
E	**Expert opinion ("There is insufficient evidence or evidence is unclear or conflicting, but this**

	is what the Work Group recommends.") Net benefit is unclear. Balance of benefits and harms cannot be determined because of no evidence, insufficient evidence, unclear evidence, or conflicting evidence, but the Work Group thought it was important to provide clinical guidance and make a recommendation. Further research is recommended in this area.
N	**No recommendation for or against ("There is insufficient evidence or evidence is unclear or conflicting.")** Net benefit is unclear. Balance of benefits and harms cannot be determined because of no evidence, insufficient evidence, unclear evidence, or conflicting evidence, and the Work Group thought no recommendation should be made. Further research is recommended in this area.

*In most cases, the strength of the recommendation should be closely aligned with the quality of the evidence;



1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 2011-2013

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Guideline data is from randomized controlled trials (RCTs), five of them, systematic reviews and meta-analyses of RCTs.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Methodology and Evidence Review: Although the Expert Panel was convened prior to the Institute of Medicine reports on practice guidelines, our evidence-based process followed most of the standards from the Institute of Medicine report, "Clinical Practice Guidelines We Can Trust" (1). The systematic review was limited to RCTs with ASCVD outcomes and systematic reviews and meta-analyses of RCTs with ASCVD outcomes. Observational studies and those with:

- The Expert Panel constructed CQs relevant to clinical practice.
- The Expert Panel identified (a priori) inclusion/exclusion (I/E) criteria for each CQ.
- An independent contractor developed a literature search strategy, based on I/E criteria, for each CQ.
- An independent contractor executed a systematic electronic search of the published literature from relevant bibliographic databases for each CQ. The date for the overall literature search was from January 1, 1995 through December 1, 2009. However, RCTs with hard ASCVD outcomes of MI, stroke, and cardiovascular death published after that date were eligible for consideration until the Expert Panel began deliberations on relevant recommendations.
- RCTs that met the inclusion criteria and were independently graded as fair or good quality were included in the evidence tables for the consideration of the Expert Panel. RCTs that were graded as poor quality were excluded.
- With the assistance of independent methodologists, this evidence base was used to develop a series of evidence statements graded on the level of the evidence (high, medium, or low).
- The Expert Panel then synthesized the evidence statements into treatment recommendations/summaries graded as A (strong), B (moderate), C (weak), D (recommend against), E (expert), and N (no recommendation).
- The final evidence statements and treatment recommendations were approved by at least a majority of voting members of the Expert Panel.
- Performed guideline implementability appraisals, planned and coordinated by the NHLBI Implementation Work Group, to identify and address barriers to guideline implementation.

Overview of the Guidelines: The RCTs identified in the systematic evidence review indicated a consistent reduction in ASCVD events from 3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors (statins) therapy in secondary and primary prevention populations, with the exception of no ASCVD event reduction in those with New York Heart Association (NYHA) class II-IV heart failure or receiving maintenance hemodialysis. The RCTs either compared fixed doses of statins with placebo or untreated controls, or compared fixed doses of higher-intensity statins with moderate-intensity statins. These trials were not designed to evaluate the effect of titrated (dose-adjusted) statin treatment to achieve pre-specified LDL–C or non-HDL–C goals.
Therefore, the Expert Panel was unable to find RCT evidence to support titrating cholesterol lowering drug therapy to achieve target LDL–C or non-HDL-C levels, as recommended by ATP III . However, the Expert Panel did find RCT evidence that use of therapy (e.g., niacin) to additionally lower non-HDL–C, once an LDL–C target was achieved, did not further reduce ASCVD outcomes. However, the Expert Panel did find extensive RCT evidence that the appropriate intensity of statin therapy should be used to reduce ASCVD risk in those most likely to benefit. The work of the Expert Panel was informed by the report of the Lifestyle and Risk Assessment Work Groups.

Evidence Statement (ES)	Level of Evidence	Rec(s)/ Section	References

In adults with CHD or acute coronary syndromes, more intensive-dose statin therapy reduced LDL–C to a greater degree (by 20 mg/dL or an additional 20%) than less intensive-dose statin therapy or placebo and produced a greater reduction in CVD events. Each 1 mmol/L (38.7 mg/dL) reduction in LDL–C reduced the RR for CVD events by approximately 28%. See Table 4 for definition of high-, moderate-, and low-intensity statin therapy. More intensive statin therapy = atorvastatin 80 mg, simvastatin 80 mg.	H	Secondary Prevention	CTT 2010(20)—data from 5 trials TNT(46) IDEAL(47) PROVE-IT(48) A–Z(119) SEARCH (128) (not included in CQ1)
Less intensive statin therapy = atorvastatin 10 mg, pravastatin 40 mg or simvastatin 20–40 mg.			
In trials of more intensive statin therapy (atorvastatin 80 mg, simvastatin 80 mg) compared with less intensive statin therapy (atorvastatin 10 mg, pravastatin 40 mg, or simvastatin 20–40 mg), women with CHD or acute coronary syndromes experienced a similar (approximately 25%) magnitude of relative CVD reduction as men (approximately 29%). Women also experienced a similar magnitude of absolute risk reduction as men	H	Secondary Prevention (women included)	CTT 2010(20)— 5 trials TNT(46) IDEAL(47) PROVE-IT(48) A–Z(119) SEARCH (128) (not included in CQ1)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The findings support the use of statins to prevent both nonfatal and fatal ASCVD events. Such an approach can reduce the large burden of disability from nonfatal stroke (for which women are at higher risk than men) and nonfatal CHD events. Primary and secondary prevention of ASCVD with statins can positively impact rising healthcare costs. In addition, a high level of evidence was found that statins reduce total mortality in individuals with a history of prior ASCVD events (e.g., secondary prevention settings). In individuals with no prior history of ASCVD events (e.g., primary prevention setting), there is moderate evidence that statins reduce total mortality in individuals at increased ASCVD risk. It should be noted, 2 meta-analyses published after the completion of the Expert Panel's systematic review provide strong evidence that statins reduce total mortality in primary prevention.

The Expert Panel defines the intensity of statin therapy on the basis of the average expected LDL–C response to a specific statin and dose. "High-intensity," "moderate-intensity," and "lower-intensity" statin therapy definitions were derived from the systematic reviews for CQ1 and CQ2. The basis for differentiation among specific statins and doses arose from the RCTs included in CQ1, where there was a high level of evidence that high-intensity statin therapy with atorvastatin 40 mg to 80 mg reduced ASCVD risk more than moderate-intensity statin therapy with atorvastatin 10 mg, pravastatin 40 mg, or simvastatin 20 mg to 40 mg bid. Classifying specific statins and doses by the percent reduction in LDL–C level is based on evidence that the relative reduction in ASCVD risk from statin therapy is related to the degree by which LDL–C is lowered. However, no variation in the relative reduction in ASCVD risk was observed after the data were adjusted for LDL–C reduction. Furthermore, there is no differentiation between the specific statins and doses used in primary and secondary prevention RCTs, based on a high level of evidence that statins reduce ASCVD risk similarly in both populations.

Percent reductions in LDL–C for a specific statin and dose were calculated for the RCTs included in individual meta-analyses conducted by the Cholesterol Treatment Trialists (CTT) in 2010 (20) in which statin therapy reduced ASCVD events. High-intensity statin therapy on average lowers LDL–C by approximately ≥50%, moderate-intensity statin therapy lowers LDL–C by approximately 30% to <30%

Table 5. High- Moderate- and Low-Intensity Statin Therapy (Used in the RCTs reviewed by the Expert Panel)*

High-Intensity Statin Therapy	Moderate-Intensity Statin Therapy	Low-Intensity Statin Therapy
Daily dose lowers LDL–C on average, by approximately ≥50%	Daily dose lowers LDL–C on average, by approximately 30% to <50%	Daily dose lowers LDL–C on average, by <30%
Atorvastatin (40†)–80 mg **Rosuvastatin 20** *(40)* **mg**	**Atorvastatin 10** *(20)* **mg** **Rosuvastatin** (*5*) **10 mg** **Simvastatin 20–40 mg‡** **Pravastatin 40** *(80)* **mg** **Lovastatin 40 mg** *Fluvastatin XL 80 mg* **Fluvastatin 40 mg bid** *Pitavastatin 2–4 mg*	*Simvastatin 10 mg* **Pravastatin 10–20 mg** **Lovastatin 20 mg** *Fluvastatin 20–40 mg* *Pitavastatin 1 mg*

Specific statins and doses are noted in bold that were evaluated in RCTs (17,18,46-48,64-67,69-78) included in CQ1, CQ2 and the CTT 2010 meta-analysis included in CQ3 (20). All of these RCTs demonstrated a reduction in major cardiovascular events. Statins and doses that are approved by the U.S. FDA but were not tested in the RCTs reviewed are listed in *italics*.

*Individual responses to statin therapy varied in the RCTs and should be expected to vary in clinical practice. There might be a biologic basis for a less-than-average response.

†Evidence from 1 RCT only: down-titration if unable to tolerate atorvastatin 80 mg in IDEAL (47).

‡Although simvastatin 80 mg was evaluated in RCTs, initiation of simvastatin 80 mg or titration to 80 mg is not recommended by the FDA due to the increased risk of myopathy, including rhabdomyolysis.

bid indicates twice daily; FDA, Food and Drug Administration; IDEAL, Incremental Decrease through Aggressive Lipid Lowering study; LDL–C, low-density lipoprotein cholesterol; and RCTs, randomized controlled trials.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

No real harms but potential for patients with a low tolerance or intolerance to statins:

Adherence to lifestyle and to statin therapy should be re-emphasized before the addition of a nonstatin drug is considered (Figure 5). RCTs evaluating the ASCVD event reductions from nonstatins used as monotherapy were reviewed as well as RCTs evaluating the additional reduction in ASCVD events from nonstatin therapy added to statin therapy. The panel could find no data supporting the routine use of nonstatin drugs combined with statin therapy to reduce further ASCVD events. In addition, identification of any RCTs that assessed ASCVD outcomes in statin-intolerant patients was not found.

Clinicians treating high-risk patients who have a less-than-anticipated response to statins, who are unable to tolerate a less-than-recommended intensity of a statin, or who are completely statin intolerant may consider the addition of a nonstatin cholesterol-lowering therapy. High-risk individuals include those with ASCVD, those with LDL–C ≥190 mg/dL and individuals with diabetes. In this situation, this guideline recommends clinicians preferentially prescribe drugs that have been shown in RCTs to provide ASCVD risk-reduction benefits that outweigh the potential for adverse effects drug-drug interaction, and patient preferences.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
Template_MeasSubm_Evidence_062915_WCHQ.pdf

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., the benefits or improvements in quality envisioned by use of this measure)*
Is a composite measure - all or none.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Level of Analysis: Clinic
Number of Clinics: 121
Minimum Number of Patients/Clinic: 100
Minimum Number of Providers/Clinic: 3
Number of Patients: 42,290
Number of Medical Groups Represented: 14
Measurement Period: January 1, 2014 – December 31, 2014

Summary Statistics
Mean: .5862
Median: .5868
Standard Deviation: 0.0712
Min: .379
Max: .750

Percentile	
Min	.379
10th	.485
20th	.529

30th	.557
40th	.572
50th (Median)	.587
60th	.603
70th	.627
80th	.645
90th	.672
Max	.750

Distribution of Scores

	Count	Percent
.379 to .399	3	2.5%
.400 to .499	12	9.9%
.500 to .599	55	45.5%
.600 to .699	46	38.0%
.700 to .750	5	4.1%

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
N/A -see above 1b.2

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
At this time WCHQ is not reporting this measure using disparities data. We do collect some race/ethnicity data, payer data, gender and age within our data repository so this could provided in the future.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A - see 1b.4 above.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
The Wisconsin Department of Health Services publication "The Epidemic of Chronic Disease in Wisconsin" indicates that heart disease is one of the leading causes of mortality in the state of Wisconsin ranking with Cancer at 24%. It is also the 4th highest annual Medicaid cost out of six chronic disease.

Prevention of Chronic Diseases is Powerful and Cost-effective
Given the huge economic impact of chronic diseases it is not surprising that their prevention yields a remarkable return on investment (ROI), as documented by the Trust for America's Health (2008). In Wisconsin, adequately funded community-based programs that address insufficient physical activity, unhealthy diet, and tobacco use would yield a return of $6.20 for every $1 spent over the course of five years, with a potential annual savings of $338 million. Remarkably, even such a substantial return can be considered an underestimate since it does not include gains in worker productivity, reduced absenteeism at work and school, and

enhanced quality of life. Despite these convincing numbers, the vast majority of health care spending in the United States, as much as 95 percent, is directed toward medical care and biomedical research and not on prevention (Institute of Medicine, 2003). The recent passage of national health care reform (The Patient Protection and Affordable Care Act) however, focuses on the importance of prevention as a means to reduce future health care costs.
Promoting healthy environments in communities is smart policy, because they enable community residents to live healthier lives. Studies increasingly suggest that businesses are also likely to benefit. For example, relatively low-cost environmental changes in the workplace that resulted in a 5% weight loss for overweight or obese employees would reduce annual medical/absenteeism costs by about $90 per person (Trogdon, et al., 2009).

Additionally a second publication titled "Wisconsin Heart Disease and Stroke Surveillance Summary Update" January 2007 by the Wisconsin Cardiovascular Health Program, the Wisconsin Department of Health and Family Services, the Division of Public Health and the Bureau of Community Health Promotion indicates that cardiovascular disease (CVD) is the leading cause of death and disability in the United States and in the state of Wisconsin. In 2004 there were 96,000 hospitalizations due to cardiovascular related illness in Wisconsin resulting in over 2.5 billion in associated charges. Death and disability from CVD can be reduced by modifiable risk factors such as quitting smoking and lowering Blood Pressure, both component measures in this all or none measure.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
The Wisconsin Department of Health Services publication "The Epidemic of Chronic Disease in Wisconsin" P-00238 (12/10)

Wisconsin Heart Disease and Stroke Surveillance Summary Update" January 2007 by the Wisconsin Cardiovascular Health Program, the Wisconsin Department of Health and Family Services, the Division of Public Health and the Bureau of Community Health Promotion

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

1d. Composite Quality Construct and Rationale

1d.1. **A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
  - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient); or
  - any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient).

1d.1. Please identify the composite measure construction: two or more individual performance measure scores combined into one score

1d.2. **Describe the quality construct, including:**

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

In November 2013, the ACC and AHA Task Force on Practice Guidelines released updated guidance for the treatment of blood cholesterol. The new recommendations remove treatment targets for LDL-C for the primary or secondary prevention of atherosclerotic cardiovascular disease (ASCVD) and recommend high or moderate intensity statin therapy based on patient risk factors. Four major statin benefit groups were identified for whom ASCVD risk clearly outweighs the risk of adverse events. Individuals with ASCVD are one of the identified groups. Based on these guidelines changes the WCHQ Measurement Advisory Committee made the decision to no longer publicly report LDL Testing and Control measures for this population and to replace these measures with a measure of Statin Use with. Simultaneously the decision was made to publicly report a measure if IVD All or None Optimal Control. These decisions were made with guidance from the WCHQ Measure Selection Policy and associated criteria. The

All-Or-None method is a more complete way of reporting the IVD measure and has multiple goals. All goals must be reached by each patient in order to meet this intermediate clinical outcome measure. This method was chosen because of the benefits it provides to both the patient and the practitioner. First, this methodology more closely reflects the interests and likely desires of the patient. With the data collected in a single score, patients can easily look and see how their provider group is performing on these criteria rather than trying to make sense of multiple scores on individual measures. Second, this method represents a systems perspective emphasizing the importance of optimal care through a patient´s entire healthcare experience. Third, this method gives a more sensitive scale for improvement. Whether reported at the organization, clinic site or provider level, for those scoring high marks on individual measures, the All-or-None measure will give room for benchmarks and additional improvements to be made.

This measure contains four goals. All four goals within a measure must be reached in order to meet that measure. The numerator for the all-or-none measure should be collected from the organization´s total IVD denominator. Using the IVD denominator optimal results include: 1) Most recent blood pressure measurement is less than 140/90 mm Hg -- And 2) Most recent tobacco status is Tobacco Free -- And 3) Daily Aspirin or Other Antiplatelet Unless Contraindicated -- And 4) Statin Use.

The four individual measures and resulting composite are based on current primary prevention guidelines from the American Heart Association (AHA) and the American College of Cardiology (ACC). Evidence from clinical trials supports and broadens the merits of risk-reduction therapies for patients with established coronary and other atherosclerotic vascular disease, including peripheral arterial disease, atherosclerotic aortic disease, and carotid artery disease.

1d.3. **Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.**

Also indicated in 1d.2 above, this method was chosen because of the benefits it provides to both the patient and the practitioner. First, this methodology more closely reflects the interests and likely desires of the patient. With the data collected in a composite score, patients can easily look and see how their provider group is performing on these criteria rather than trying to make sense of multiple scores on individual measures. Second, this method represents a systems perspective emphasizing the importance of optimal care through a patient's entire healthcare experience. Third, this method gives a more sensitive scale for improvement. Whether reported at the organization, clinic site or provider level, for those scoring high marks on individual measures, the All-or-None measure will give room for benchmarks and additional improvements to be made.

Support for All-or-None measurement referenced by Nolan T, Berwick DM. All-or-None Measurement Raises the Bar on Performance. JAMA. 2006 Mar 8;295(10):1168-70.

1d.4. **Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**

The composite measure result is based off of the four individual measure numerators that construct the all or none measure. The numerator for the all-or-none measure is collected from the organization, clinic site or provider´s total IVD denominator and the patient must meet all four individual measures to be numerator compliant for the all or none measure. All individual component scores are given equal weighting when combined into the composite. The measure can be reported through an all electronic data collection method-based on entire IVD Denominator, a hybrid method-based on Administrative Review Denominator and Manual Review Sample, or a random sample method-based on Sample Population. The hybrid or random sample method are used if there will be missing data elements. The sample size for for chart review is determined by the WCHQ Sample Size Calculator, available at the following link: http://www.wchq.org/calculator/index.php.

NATIONAL QUALITY FORUM—Composite Measure Testing (subcriteria 2a2, 2b2-2b7, 2d)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Composite Measure Title: Ischemic Vascular Disease Care: All or None Outcome Measure-Optimal Control
Date of Submission: 6/29/2015
Composite Construction:

☐Two or more individual performance measure scores combined into one score
☒ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)
☐ Any-or-none measures (e.g., any or none of a list of adverse outcomes experienced, or inappropriate or unnecessary care processes received, by each patient)

Instructions: Please contact NQF staff before you begin.

- If a component measure is submitted as an individual performance measure, the non-composite measure testing form must also be completed and attached to the individual measure submission.
- 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 composite measures, sections 1, 2a2, 2b2, 2b3, 2b5, and 2d must be completed.**
- **For composites with outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitions** (e.g., claims and EHRs), section **2b6** 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), validity (2b2-2b6), and composites (2d) 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.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

2d. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2d1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2d2.the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(*if not conducted or results not adequate, justification must be submitted and accepted)*

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.

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 different components in the composite, indicate the component after the checkbox.*)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record

☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/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*).

Background of WCHQ's Reporting Activities: WCHQ is a voluntary, non-profit (501c3) consortium of organizations committed to using the public reporting of comparative measures of performance to catalyze improvements in the quality and affordability of healthcare, as well as the health status of individuals and communities, in Wisconsin. In addition, WCHQ designs and facilitates collaborative learning sessions to promote the active sharing of "best practices" in an effort to elevate and accelerate improvement across the state. WCHQ has received national recognition for its work and has made numerous contributions to the emerging evidence base on the science of measurement and reporting. Equally important, WCHQ and the healthcare provider organizations within its membership have directly contributed to a significant improvement in the quality of care, as reflected in the measures reported via the WCHQ website (www.wchq.org). Peer-reviewed research has shown that WCHQ's public reporting has played a significant role in catalyzing quality improvement among the physician groups in Wisconsin. WCHQ has a total of 38 members organizations, representing approximately 65% of the physicians licensed to practice in Wisconsin.

Dataset Background: An existing data set was used. WCHQ members submit data in one of two ways as follows:

Repository Based Submission (RBS):

- WCHQ members submit global patient level files of patient demographic, encounter, and clinical (labs, other tests, tobacco related elements, blood pressure dates and results, medications) data into the data repository through a secure, HIPAA compliant portal. See the attached RBS File Formats document for complete documentation on data elements and file formats in Appendix A.
- Certain data elements that are unique to a given organization are mapped within the data tool to be recognized as a standard data element. Data can be normalized prior to upload to ensure that data is clean in certain cases, i.e. a lab test should be the final result.
- The RBS tool's centrally programmed measure specifications calculate performance results for internal use by the member organization, for purposes of PQRS reporting to CMS, and for reporting on WCHQ's public website- http://www.wchq.org/reporting/results.php?category_id=0&topic_id=27&source_id=0&providerType=0&region=0&measure_id=205

Alternative Data Submission Method:

- WCHQ members use detailed measure specifications to program measures internally and submit aggregate denominator and numerator data to WCHQ's web-based reporting tool for reporting results on WCHQ's public website. De-identified patient level data is additionally submitted for validation purposes.

Testing for the IVD Care All or None Optimal Control measure involved the following:

- 50,758 patients in the population denominator at the group level
- 17 WCHQ member organizations (group level) publicly reported this measure
- 121 member clinic sites reported on 42,290 patients. This number is smaller than the group level patient number because at the clinic site level there must be a minimum of 100 patients in the denominator at each clinic in order to report to avoid variance in results that could occur with a smaller minimum.

15 groups reported all electronically (Total Population) and 2 groups reported using the random sample methodology.

1.3. What are the dates of the data used in testing? 01/01/2013-12/31/2014

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

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

- 50,758 patients in the population denominator at the group level, including:
  - Male and Female
  - Ages 18-75
  - All races
  - All payers
  - Diagnosis of IVD
  - Alive as of the last day of the Measurement Period
- 17 WCHQ member organizations (group level) publicly reported this measure
- 121 member clinic sites reported on 42,290 patients. This number is smaller than the group level patient number because at the clinic site level there must be a minimum of 100 patients in the denominator at each clinic in order to report in order for the result to be considered statistically significant.

Includes all primary care providers employed by the member organization. This measure can include cardiologists as a "Measure Specific Specialist", if the organization chooses to include them and typically if they are included one of the visits that counts towards the denominator would be with a PCP.

All patients 50,578 are included in the numerator testing. Because the data base is global, all patients that reside in the data repository are initially included and many gradually fall out by not meeting the specific denominator criteria (see below):

[Question 1] – **Is this a patient with the disease, or condition?**

CORONARY ARTERY DISEASE (OR CAD RISK EQUIVALENT) DIAGNOSIS RELATED OUTPATIENT VISITS

Those patients with a total of two or more visits during the last 24 months [Measurement Period + Prior Year] from Table IVD-4 (Office Visit Encounter Codes-Outpatient) with any provider (MD, DO, PA, NP) within the Physician Group on different dates of service coded *(including primary and secondary diagnoses)* with diagnosis codes from Table IVD-1 (Coronary Artery Disease) or Table IVD-2 (CAD Risk-Equivalent Conditions). The following criteria apply:

Any combination of two or more diagnosis codes from either Table IVD-1 or Table IVD-2, on different dates of service.

OR

ACUTE CORONARY EVENT- RELATED HOSPITAL VISITS Those patients who had a minimum of *one* hospital related visit (excluding Emergency and Lab Only visits) for an Acute Coronary Event from Table IVD-3 during the last 24 Months [Measurement Period + Prior Year].

[Question 2] – **Is this a patient whose care is managed within the physician group?**

Those patients who have at least two Primary Care Office Visit (Table IVD-4) in an ambulatory setting, regardless of diagnosis code, on different dates of service, to a PCP or Cardiologist in the past 24 months [Measurement Period + Prior Year]. If Cardiologist is not considered a PCP, at least one of the two office visits must be to a PCP.

[Question 3] – **Is this a patient current in our system?**

Those patients who had at least *one* Primary Care Office Visit (Table IVD-4) in an ambulatory setting, regardless of diagnosis code, with a PCP or a Cardiologist during the last 12 Months [Measurement Period].

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.

There are no differences in the data for different aspects of testing.

2a2. RELIABILITY TESTING

2a2.1. What level of reliability testing was conducted?

Note*: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*

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

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

The WCHQ staff used the methodology outlined by John L. Adams, Ph.D in his tutorial, "Reliability in Provider Profiling". Dr. Adams' methodology was applied to the clinic-level performance of the IVD: All – or-None

measure for clinics with greater than 100 patients in the denominator, and greater than 2 providers attributed to the clinic site, which are the public reporting parameters established by WCHQ.

The methodology outlined by Adams' and used to calculate the average reliability of a WCHQ clinic performance score involves several steps, including:

1. The calculation of clinic-to-clinic variance, using the following process:
    a. Variance equation: $\sigma^2_{\text{clinic-to-clinic}} = (\alpha\ \beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$
    b. The α and β were calculated using the publicly available beta binomial SAS macro, BETABIN, developed by Qi Statistics
    c. $\alpha = 35.6267$, $\beta = 25.0007$, $\sigma^2_{\text{clinic-to-clinic}} = 0.00393$

2. The reliability of each individual clinic, using the equation, Reliability = $\sigma^2 / (\sigma^2 + (p(1 - p)/n))$
    a. p = clinic performance rate, n = number of patients in the measure denominator

3. The reliability for each clinic was then calculated. The summary statistics are below:
    a. n = 121 clinics
    b. Average reliability = .7817
    c. Median reliability = .7739
    d. Max = .9911, Min = .6159

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

Average Reliability = 0.7817

Below is the output of the BETABIN macro.

IVD All or None: Raw Data Summary

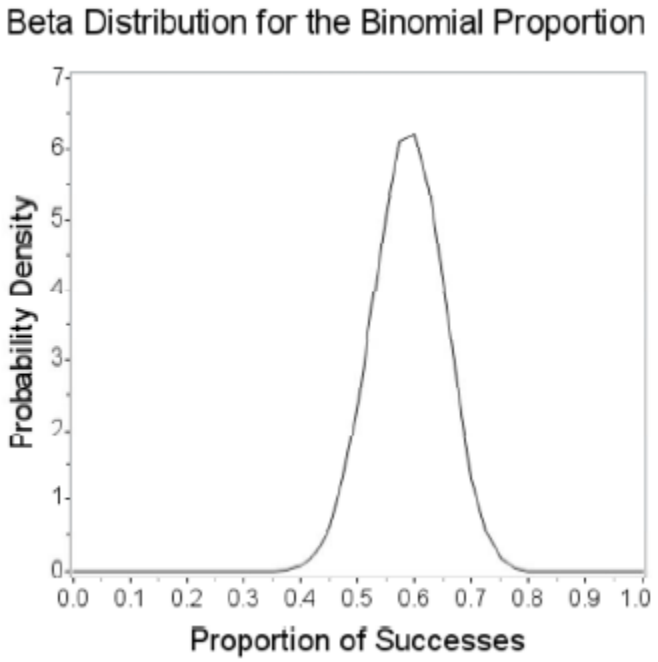
Item	Value
Data Set	WORK.WCHQII
Response Variable (Events)	NumRound
Response Variable (Trials)	Population_Denominator
Number of Observations Read	121
Number of Observations Used	121
Number of Events	24991
Number of Trials	42290

IVD All or None: Simple Binomial Model

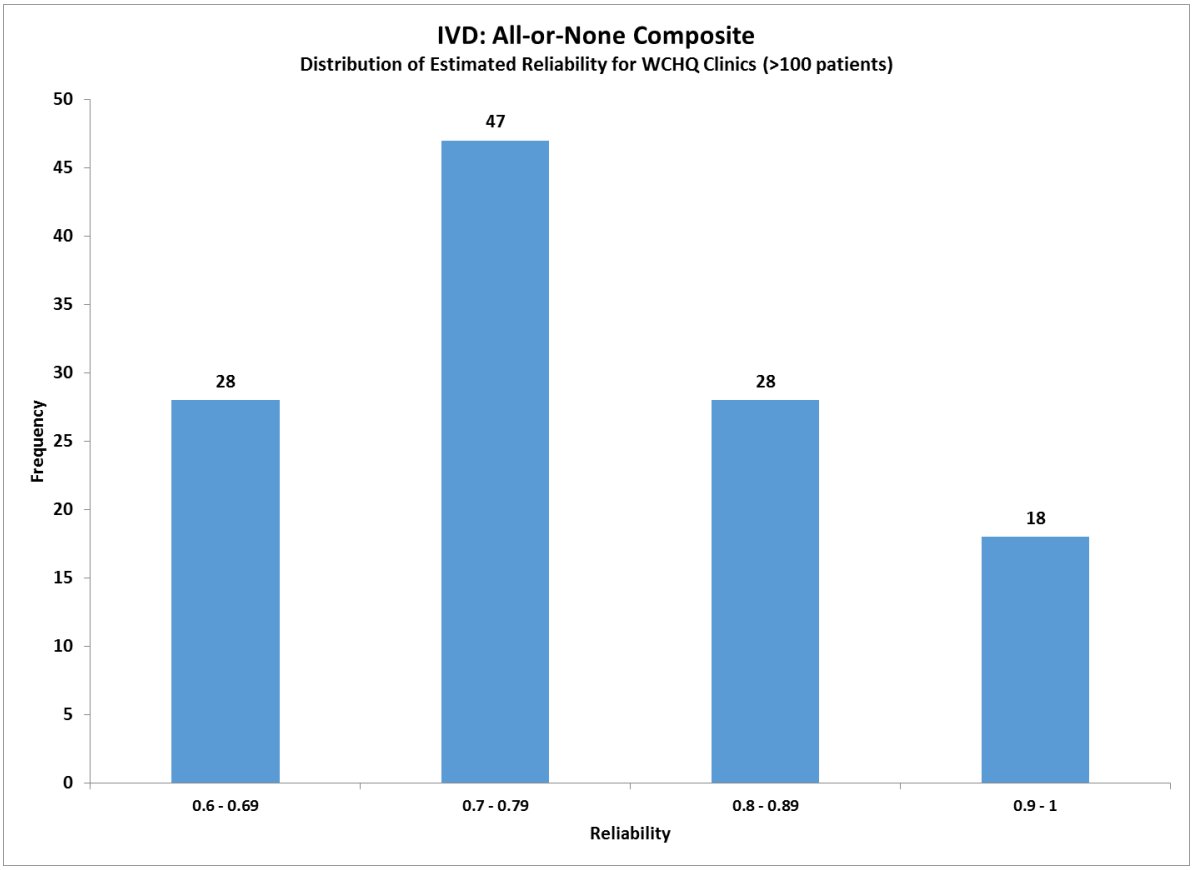
parameter	Estimate	Standard Error	t Value	Prob > |t|	Alpha	Lower	Upper
mu	0.5909	0.0024	247.17	<.0001	0.05	0.5863	0.5956
|mu-0.5|	0.0909	0.0024	38.04	<.0001	0.05	0.0863	0.0956

IVD All or None: Beta-Binomial Model Parameters

parameter	Estimate	Standard Error	t Value	Pr > |t|	Alpha	Lower	Upper
mu	0.5876	0.006453	91.07	<.0001	0.05	0.5750	0.6003
alpha	35.6267	5.7659	6.18	<.0001	0.05	24.3254	46.9281
beta	25.0007	4.0170	6.22	<.0001	0.05	17.1273	32.8741
gamma	0.01623	0.002567	6.32	<.0001	0.05	0.01119	0.02126
theta	0.01649	0.002653	6.22	<.0001	0.05	0.01129	0.02169
|mu-0.5|	0.08763	0.006453	13.58	<.0001	0.05	0.07499	0.1003



Below is a chart outlining the reliability distribution for all of the clinics included in this analysis.



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?*)

It is generally accepted that a reliability score of greater than 0.70 is sufficient for determining performance differences between groups. The average level of reliability for the WCHQ member clinics is 0.7817, above the minimal reliability threshold. In fact, of those clinics used in this sample, 77% of them have a reliability score greater than .70, and zero clinics below .60. This analysis demonstrates that this IVD: All-or-None measure has a sufficient level of average reliability to determine performance differences between clinics.

2b2. VALIDITY TESTING

Note*: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

2b2.1. What level of validity testing was conducted?

☐ **Composite performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

☒ **Systematic assessment of content validity**

☐ **Validity testing for component measures** *(check all that apply)*

Note*: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☐ **Endorsed (or submitted) as individual performance measures**

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

☐ **Empirical validity testing of the component measure score(s)**

☒ **Systematic assessment of face validity of component measure score(s) as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. 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)*

WCHQ's validation model ensures the accuracy of the publicly reported performance measures. Denominator validation is done for each reporting entity (RE) on all measures annually. The validation process also includes a 3-year numerator validation schedule, based upon the following selection criteria:

1. **Random Selection without Replacement:** Following baseline validation, each RE is randomly selected over a 3-year period to be validated for each measure numerator. Once validation has been completed for a given RE for a given numerator, it is not selected again for the same numerator during the 3-year period of time, unless an identified issue arises.
2. **Outlier Results:** Results that appear significantly higher or lower than the mean or results that vary greatly for a RE between two reporting periods will be validated.
3. **New Organizations:** In the first year that a new RE reports measures, they are validated for all measure elements (denominator and numerator) during each reporting period.
4. **New Measures**: Upon the release of a new measure, all REs are validated for each measure element (denominator and numerator).
5. **Significant Changes to Database Structure**: For any significant change to a RE's database structure, all elements (denominators and numerators) are validated for a given reporting period.

The majority of WCHQ members report through the repository based submission (RBS) outlined in section 1.2 above and as such submit global patient and encounter files and access to patient level data is available through this data source. Files needed for this measure are the Patient File, Encounter File, Clinical Data File, Blood Pressure File, Medication File and Tobacco File. All necessary data elements are documented in the RBS File Formats document (attached in Appendix A). In addition, certain data elements are cross-mapped in order to meet a standard code (i.e. if the data element is an A1C in their system they need to cross-map it within RBS to indicate that it is the equivalent of CPT code 83036. Primary Care Provider documentation is also mapped to ensure that the correct providers are included in the respective measures. Additionally, when the data files are uploaded there are various checks to ensure that all data elements that need to be cross-mapped are considered and that files are in the correct format so that there will not be missing data. All of the WCHQ quality measures and selected PQRS measures are programmed into the RBS according to the detailed measures specifications and the measures are then calculated on behalf of the reporting entity through this standard process, so there is no chance of misinterpretation of the measure specification by an individual reporting entity. The results are then available for each measure at the group level, clinic site level and the provider-patient level. Patient level results are used to verify that all data elements are cross-mapped correctly, that the right patients meet the denominator criteria (i.e. appropriate office visit codes and dates, and diagnosis if that is part of the measure)and to ensure that there is no missing data. Reporting entities have access to this patient level data and are required to randomly sample patients that didn't meet the numerator to ensure that they are not missing any data elements or sources in the data files. WCHQ also performs occasional on-site visits to assist members with data collection processes.

Four of the 19 entities that attempted to report the IVD All or None Measure used the alternate reporting method (described in section 1.2 above). These entities used the detailed measure specification to program the measure internally and report aggregate results through the WCHQ data tool portal using a group and clinic site

level reporting template. In addition these entities are required to upload patient level data files and complete on-line denominator and numerator validation forms that describe how patients met both the denominator and numerator. See "Fields Required for Measure Validation" below, which is taken from the detailed measure specification. Members are required to provide documentation outlining their PCP and Measure Specific Specialist definition and any internal data elements that map to a standard code.

FIELDS REQUIRED FOR MEASURE VALIDATION

Validation of this measure will require patient level data files for Administrative Data and/or for Manual Review. The following indicates fields needed for validation, which may be helpful to consider when querying the measure:

Denominator Data File fields:

1. Generic Patient Identifier (can be medical record number or other ID)
2. Primary Care Office Visit Dates
3. Inpatient Visit Date (if applicable)
4. Provider Specialty
5. Patient Date of Birth
6. CAD or CAD Risk-Equivalent Diagnosis Codes

Numerator Data File fields:

1. Daily Aspirin or Daily Other Antiplatelet Therapy documented as active in the medical record at any time during the measurement period with data entry including:
   - Generic Patient Identifier (can be medical record number or other ID)
   - Aspirin or other Antiplatelet medication drug name
   - Drug frequency
   - Medication status indicated as active during the measurement period
   - Contraindications (if any apply)

2. Blood Pressure Control within the last 12 months
   - Patient Identifier
   - Blood Pressure Date(s) of Service
   - Blood Pressure Result(s)

3. Most Recent Tobacco Status
   - Patient Identifier (Can be medical record number or other ID)
   - Tobacco Status
   - Encounter Date of Service Associated with Tobacco Status

4. Statin Use
   - Patient Identifier (Can be medical record number or other ID)
   - IVD Diagnosis, if applicable
   - Statin Medication name
   - Medication status indicated as active during the measurement period

Site Level Reporting fields:

- Clinic Name
- Period
- Metric ID
- Clinical Topic
- Measure
- Clinic ID
- Clinic Name
- Metric Level (for A1C and LDL Testing and Control measures)
- Payer (optional)
- Numerator
- Denominator
- Percentage

- Provider Count
- Provider Minimum Count Flag
- Patient Minimum Count Flag

Two of the 19 entities that attempted to report this measure had issues that were discovered during the validation process and the decision was made for them to be listed as "Did Not Report" for this measurement period on the public website. The issues found were related to missing medication data, data systems being combined and staff changes in areas that work directly with the WCHQ measures. Both entities have a goal to work on resolution of the issues and plan to report the measure when it is publicly reported again in November 2015. The four individual component measure results were reviewed for accuracy as well as the combined score for the all or none measure and one issue was found in a group that sampled with selection of the wrong sampling method and this was corrected and they were still able to publicly report. Seventeen entities passed validation for reporting publicly reporting the measure.

Published results at the group level ranged from 44.80% to 70.02%. At the clinic site level results ranged from 37.92% to 75.00%.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

2b2.4. 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?*)

2b3. EXCLUSIONS ANALYSIS

Note: *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*

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

2b3.1. 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*)

2b3.2. 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*)

2b3.3. 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*)

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

Note: *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used? *(check all that apply)*

☐ **Endorsed (or submitted) as individual performance measures**

☒ **No risk adjustment or stratification**
☐ **Statistical risk model**
☐ **Stratification by risk categories**
☐ **Other,** Click here to enter description

2b4.2. 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.
WCHQ does not currently risk adjust our data.. Risk adjustment of ambulatory clinical outcome and process measures is not currently done that we are aware of at the regional or national level by virtue of the type of measures that these are in comparison to hospital-based measures. In addition, WCHQ reports on all patients and all payers.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)

2b4.4. What were the statistical results of the analyses used to select risk factors?

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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?*)

***2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
Note: *Applies to the composite performance measure.*

2b5.1. 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)*

2b5.2. 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*)

2b5.3. 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?*)

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
Note: *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*
If only one set of specifications for each component, this section can be skipped.

Note*: This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)

2b6.2. 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*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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?*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS
Note: *Applies to the overall composite measure.*

2b7.1. 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*)

If any one of the four individual component measure is missing the result, i.e. patient does not have an active aspirin order the patient will remain in the denominator but will fail the numerator. If there is no documentation of tobacco status, the patient is not compliant for this measure.

2b7.2. 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*)

WCHQ receives global data files into RBS and patient level data files for validation using the alternate reporting method so any instances of missing data are found during the validation process and the determination is made at that time through verification of the reason for the missing data or result and whether or not the measure can be reported publicly. Reasons could be an unknown data source, a documentation process issue or an incorrectly mapped data element.

2b7.3. 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*)

Patients with missing data are not excluded from the all or none measure. Data elements missing from any component are counted as a numerator fail and the patient remains in the denominator. See 2b7.1 above.

2d. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

This all or none method was chosen because of the benefits it provides to both the patient and the practitioner. First, this methodology more closely reflects the interests and likely desires of the patient. With the data collected in a composite score, patients can easily look and see how their provider group is performing on these criteria rather than trying to make sense of multiple scores on individual measures. Second, this method represents a systems perspective emphasizing the importance of optimal care through a patient's entire healthcare experience. Third, this method gives a more sensitive scale for improvement. For those organizations scoring high marks on individual measures, the All-or-None measure will give room for benchmarks and additional improvements to be made.

Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA*. 2006 Mar 8;295(10):1168-70.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc*.; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

The components of this measure were selected as secondary prevention, Level A AHA/ACC recommendations that can significantly reduce the IVD patient's risk of developing additional cardiovascular conditions.

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

This all or none method was chosen because of the benefits it provides to both the patient and the practitioner. First, this methodology more closely reflects the interests and likely desires of the patient. With the data collected in a composite score, patients can easily look and see how their provider group is performing on these criteria rather than trying to make sense of multiple scores on individual measures. Second, this method represents a systems perspective emphasizing the importance of optimal care through a patient's entire healthcare experience. Third, this method gives a more sensitive scale for improvement. For those organizations scoring high marks on individual measures, the All-or-None measure will give room for benchmarks and additional improvements to be made.

Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA*. 2006 Mar 8;295(10):1168-70.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

There is no weighting of the component measures.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

N/A

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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 : Ischemic Heart Disease, Coronary Artery Disease

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

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.)*
http://onlinecommunity.wchq.org/default.asp?page=qcdr

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

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

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Measure has not been previously endorsed.

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
All-or-None Outcome Measure (Optimal Control) - Using the IVD denominator optimal results include:

- Most recent blood pressure measurement is less than 140/90 mm Hg

And

- Most recent tobacco status is Tobacco Free

NOTE: If there is No Documentation of Tobacco Status the patient is not compliant for this measure.
And

- Daily Aspirin or Other Antiplatelet Unless Contraindicated

And

- Statin Use

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Denominator: A minimum of two CAD or CAD Risk-Equivalent Condition coded office visits OR one Acute Coronary Event (AMI, PCI, CABG) from a hospital visit and must be seen by a PCP / Cardiologist for two office visits in 24 months and one office visit in 12 months.

Numerators: Twelve Month Measurement Period

S.6. **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, 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.

NOTE: All code tables and associated codes referenced in this document are included in the Excel File attached at step S2b.

• DAILY ASPIRIN OR OTHER ANTIPLATELET MEDICATIONS THERAPY UNLESS CONTRAINDICATED (Figure IVD-2) This measure assesses the percentage of patients with documentation within the medical record of daily Aspirin or daily other antiplatelet agent at any time during the measurement period demonstrated through any of the following:

1. Documentation of an active prescription for daily Aspirin (see suggested list in Table IVD-6) or daily or other antiplatelet medications (see acceptable medications in Table IVD-7)
2. Documentation on the patient's medication list of active daily usage of Aspirin (see suggested list in Table IVD-6) or daily other antiplatelet medications (see acceptable medications in Table IVD-7)
3. Contraindication to Aspirin

a. Contraindications will count as numerator compliant. Any valid contraindication date prior to the end of the measure end date will count as compliant. There is no limit on the look back date, but the date of documentation or onset date must occur prior to the end of the measurement period.

b. Accepted contraindications:

i. History of gastrointestinal (GI) bleed (see codes in Table IVD-8)

ii. History of intracranial bleed (ICB) (see codes in Table IVD-8)

iii. History of GI Bleed or ICB from an ICD-9 diagnosis-based problem list or past medical history. There is no limit on the look back date, but the date of documentation or onset date must occur prior to the end of the measurement period.

iv. Anticoagulant Use (see acceptable list of Medications in Table IVD-9). There must be documentation of an active anticoagulant at any time during the Measurement Period.

• BLOOD PRESSURE CONTROL (Figure IVD-2)

The number of patients in the denominator whose blood pressure (BP) is adequately controlled during the Measurement Period. Adequate control is a representative systolic Blood Pressure less than 140 mm Hg and a representative diastolic Blood Pressure less than 90 mm Hg.

IDENTIFYING A REPRESENTATIVE BLOOD PRESSURE

Blood Pressure Selection Criteria:

a) Blood Pressure reading must have been obtained during the Measurement Period.

b) Systolic and Diastolic numbers must be from the same BP reading.

c) A controlled BP requires that both the systolic and diastolic readings must be less than140/90.

d) Exclusions: Inpatient Stays, Emergency Room Visits, Urgent Care Visits, and Patient Self-Reported BP's (Home and Health Fair Blood Pressures)

e) Inclusions: Any office visit encounter, including Nurse Only BP Checks, not listed under Exclusions above. NOTE: A BP performed at a patient's home by a nurse who then inputs the result into an EMR counts as a Nurse Only BP.

• Select the Blood Pressure from the most recent visit.

• In the event that multiple Blood Pressures are recorded in the same day of service, select any reading that is controlled. If none are in control, select an uncontrolled reading.

• If no Blood Pressure is recorded during the Measurement Period, the patient is assumed to be "not controlled".

3. TOBACCO FREE (Figure IVD-2)

The number of patients in the denominator whose most recent tobacco documentation status with any provider within the 12 month measurement period is Tobacco Free.

Tobacco Use Definition:

• Cigarette
• Cigar
• Pipe Smoking
• Smokeless Tobacco (Chewing Tobacco, Snuff, etc.)

Tobacco Use Status can be identified by any of the following criteria:

1. Documentation stating that the patient has been asked if they are one of the following during the Measurement Period with the numerator compliant goal of Tobacco-Free:

1. Tobacco-Free (see examples below):

a. Former tobacco user

b. Never used
c. Non-tobacco user
d. Passive smoker
2. Non Tobacco-Free
a. Current tobacco user
3. No Documentation: The subset of denominator patients who did not have documentation of tobacco status during the last 12 Months [Measurement Period]
2. ICD-9, CPT, HCPCS and CPT-II Codes indicating tobacco use status during the Measurement Period) from billing or encounter data only. Do not use the problem list for these codes. (Table IVD-10)

4. STATIN USE (Figure IVD-2)
This measure assesses the percentage of patients with documentation within the medical record of statin use at any time during the measurement period demonstrated through any of the following:
1. Documentation of an active prescription for a statin (see acceptable medications in Table IVD-11)
2. Documentation on the patient's medication list of active usage of a statin (see acceptable medications in Table IVD-11)

5. ALL OR NONE OUTCOME MEASURE

IVD All-or-None Measure
The IVD All-or-None Measure is one outcome measure (optimal control). The measure contains four goals. All goals must be reached in order to meet that measure. The numerator for the all-or-none measure should be collected from the organization's total IVD denominator.

All-or-None Outcome Measure (Optimal Control) - Using the IVD denominator optimal results include:
- Most recent blood pressure measurement is less than 140/90 mm Hg

And

- Most recent tobacco status is Tobacco Free

NOTE: If there is No Documentation of Tobacco Status the patient is not compliant for this measure.
And
- Daily Aspirin or Other Antiplatelet Unless Contraindicated

And
- Statin Use

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Patients with CAD or a CAD Risk-Equivalent Condition 18-75 years of age and alive as of the last day of the MP.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk : Individuals with multiple chronic conditions

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
NOTE: All code tables and associated codes referenced in this document are included in the Excel File attached at step S2b.

Patients eligible for inclusion in the denominator include (See Figure IVD-1):
[Question 1] – Is this a patient with the disease, or condition?

CORONARY ARTERY DISEASE (OR CAD RISK EQUIVALENT) DIAGNOSIS RELATED OUTPATIENT VISITS
Those patients with a total of two or more visits during the last 24 months [Measurement Period + Prior Year] from Table IVD-4 (Office Visit Encounter Codes-Outpatient) with
any provider (MD, DO, PA, NP) within the Physician Group on different dates of service coded (including primary and secondary diagnoses) with diagnosis codes from Table
IVD-1 (Coronary Artery Disease) or Table IVD-2 (CAD Risk-Equivalent Conditions). The following criteria apply:

Any combination of two or more diagnosis codes from either Table IVD-1 or Table IVD-2, on different dates of service.

OR

ACUTE CORONARY EVENT- RELATED HOSPITAL VISITS
Those patients who had a minimum of one hospital related visit (excluding Emergency and Lab Only visits) for an Acute Coronary Event from Table IVD-3 during the last 24 Months [Measurement Period + Prior Year].

[Question 2] – Is this a patient whose care is managed within the physician group?
Those patients who have at least two Primary Care Office Visit (Table IVD-4) in an ambulatory setting, regardless of diagnosis code, on different dates of service, to a PCP or Cardiologist in the past 24 months [Measurement Period + Prior Year]. If Cardiologist is not considered a PCP, at least one of the two office visits must be to a PCP.
[Question 3] – Is this a patient current in our system?
Those patients who had at least one Primary Care Office Visit (Table IVD-4) in an ambulatory setting, regardless of diagnosis code, with a PCP or a Cardiologist during the last 12 Months [Measurement Period].

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
There are no denominator exclusions

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
N/A

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
This measure could be stratified by payer and this is documented in Appendix A of the measure specification, however, WCHQ does not currently publicly report the measure in a stratified manner.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Other (specify):
If other: Percentage

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk*

adjustment; etc.)
NOTE: Flow diagrams outlining the measure logic are included in step S.19.below at A.1 and is also included in the measure specification on pages 4 and 8 available at the URL identified in S.1.

The denominator algorithm is applied by identifying the target population based on codes and appropriate office visits during the designated timeframe. Once the denominator population has been identified the numerator logic is applied to all patients in the denominator to determine which patients meet each individual numerator and for the All or None measure which patients meet all four numerators for the timeframe.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
If a practice is unable to obtain the data in an all electronic format, sampling is allowed via either hybrid or random sampling.

RANDOM SAMPLE METHODOLOGY:
- Population Denominator (N) (CAD or CAD Risk-Equivalent patients 18-75 years of age)
? Population Sample (n) (r) (Patients in Denominator Population whose records will be reviewed)
o (n)=Population Sample and (r)=Patients Reviewed equal the same number
o The Population Sample size must be determined using the WCHQ Sample Calculator http://www.wchq.org/calculator/index.php
- Numerators

1. Daily Aspirin or daily other Antiplatelet Medication Therapy unless contraindicated documented in the medical record as active at any time during the measurement period
2. Most recent Blood pressure controlled to a level of less than 140/90 mm Hg
3. Most recent Tobacco Status
4. Statin Use
5. All or None Optimal Control

Upon entry of these numbers, the rate is automatically calculated

HYBRID METHODOLOGY:
- Population Denominator (N) (CAD or CAD Risk-Equivalent patients 18-75 years of age)
? Administrative Review Denominator (Patients in Total Denominator Population whose numerator information is obtained through administrative data)

? Administrative Review Numerators

1. Daily Aspirin or daily other Antiplatelet Medication Therapy unless contraindicated documented in the medical record as active at any time during the measurement period
2. Most recent Blood pressure controlled to a level of less than 140/90 mm Hg
3. Most recent Tobacco Status
4. Statin Use
5. All or None Optimal Control

? Manual Review Denominator (Patients in Total Denominator Population whose numerator information cannot be obtained through administrative data)
? Manual Review Sample Size (Patients in Manual Review Denominator Population whose records will be reviewed)
o The Manual Review Sample size must be determined using the WCHQ Sample Calculator plus a 10% over sample http://www.wchq.org/calculator/index.php
- Manual Review Numerators

1. Daily Aspirin or daily other Antiplatelet Medication Therapy unless contraindicated documented in the medical record as active at any time during the measurement period
2. Most recent Blood pressure controlled to a level of less than 140/90 mm Hg
3. Most recent Tobacco Status
4. Statin Use
5. All or None Optimal Control

Upon entry of these numbers for each numerator, the Rates, Weight Factors and Total Reviewed are automatically calculated. Total Reviewed equals Administrative Review Denominator + Manual Review Sample Size.

These instructions are also included in the specification identified in S.1.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
When possible data is collected in an all electronic format and will be all-inclusive. A patient level validation and verification process that involves comparing results at the patient level to the practice EMR data source is performed to assist with finding any missing data. If practices are aware that they will have gaps in the electronic data that is available, the numerator can be obtained through sampling, as described in S.20 above.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Data is obtained via data extracts (.csv files) from the practice and then uploaded into the WCHQ Repository Based Submission (RBS) database. Primary files consist of a Patient File, Encounter File, Problem List File, Clinical Data File, Tobacco File, Blood Pressure File and a Medication File. Certain data elements are cross-mapped to identify internal codes. The data is then calculated for the measure and is available with results at the group, clinic site and provider level. There is documentation provided describing the process of data submission and creation of the data files. This documentation is attached at A.1.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **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.)*

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Template_MeasSubm_CompositeMeasTesting_2013-08-20-WCHQ_IVD_All_or_None-635711968805327166.docx

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)*
ALL data elements are in defined fields in a combination of electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

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.**
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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
WCHQ has been collecting data for reporting clinical quality measures since 2004. Over this time we have refined our data submission and measure calculation process through use of our Repository Based Submission (RBS) system which allows us to have a large data repository of claims and clinical data. Use of RBS has streamlined the measurement and reporting process for our reporting entities by allowing them to upload and map their data and not have to build complex measure queries for multiple measures. The programming of each measure is done by WCHQ´s Technology Vendor, Ancilla Partners and this in turn ensures the measure is always calculated in a standard format when reporting via RBS. In addition, the data set can be used for multiple measure initiatives. The validation process is also streamlined because the data is at the patient level and readily accessible. Reporting entities can also use the RBS data tool to run their measures at a more frequent time-frame and produce group, clinic site,and provider level reports for internal quality improvement work. The data is uploaded initially into a secure file transfer protocol site and from there is is uploaded into the RBS database. The data base is HIPAA compliant and has been audited. In addition, WCHQ members sign a Business Associated Agreement that allows them to provide this data to WCHQ for measurement reporting.

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)*.
There are no fees associated with reporting this measure. Results for WCHQ members are available on the public website, www.wchq.org. There is resource time involved in preparing the data files, cross-mapping data elements and verifying results.

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.

Planned	Current Use (for current use provide URL)
	Public Reporting Wisconsin Collaborative for Health Care Qualifty http://www.wchq.org/reporting/results.php?site_level_flag=0&measure_id=205 Payment Program Qualified Clinical Data Registry (QCDR) approved by CMS http://onlinecommunity.wchq.org/?page=qcdr Quality Improvement with Benchmarking (external benchmarking to multiple organizations) WCHQ Measures Summary http://www.wchq.org/reporting/wchq_measures_summary.php

4a.1. For each CURRENT use, checked above, provide:

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

a. Public Reporting: The Wisconsin Collaborative for Healthcare Quality (WCHQ) publicly reports and brings meaning to performance measurement information that improves the quality and affordability of healthcare in Wisconsin, in turn improving the health of individuals and communities.WCHQ is a multi-stakeholder, voluntary consortium of Wisconsin organizations. WCHQ draws its membership from health systems, medical groups, hospitals and health plans. This diverse and dynamic group includes the state´s largest health systems, Aurora Health Care and the University of Wisconsin Hospital and Clinics / University of Wisconsin Medical Foundation. The geographic area included in public reporting is statewide and also includes some patients from neighboring states, Minnesota, Iowa, Michigan and Illinois. There are 38 member organizations but not all currently publicly report this measure yet. For 2014, the first time this measure was reported, there were 17 member organization reporting at the group level and 121 at the clinic site level. There are 50,758 patients reported for this measure.

c. Payment Program: WCHQ is a Centers for Medicare and Medicaid approved Qualified Clinical Data Registry (QCDR). This is a vetted, non-PQRS measure that has been approved by CMS to report on behalf of an eligible provider for purposes of meeting PQRS reporting requirements. Because the measure was not available as a PQRS measure for reporting through the QCDR for the 2014 measurement period but will be for the 2015 PQRS reporting period. The geographic area could potentially be across the United States and and the number of entities and patients is not known at this time. The measure could also be publicly reported on Physician Compare in 2016.

f. Quality Improvement with Benchmarking: On the wchq public website there is a Measure Summary display where members can select a reporting organization and then look at their measure results. Measure results can be viewed by Top Performer, 95th, 90th, 75th and 50th percentiles, and by Average. The geographic area included in public reporting is statewide and also includes some patients from neighboring states, Minnesota, Iowa, Michigan and Illinois. There are 38 member organizations but not all currently publicly report this measure yet. For 2014, the first time this measure was reported, there were 17 member organization reporting at the group level and 121 at the clinic site level. There are 50,758 patients reported for this measure.

4a.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?*)

4a.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.*)

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

2014 is the first year this measure was published so there is no progress on improvement at this time. We will report the measure again in November 2015 and this can be measured at that time.

4b.2. **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.**
As stated in 4b.1 above, 2014 is the first year this measure was published so there is no progress on improvement at this time. We will report the measure again in November 2015 and improvement will be measured at that time.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
None.

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)**
0076 : Optimal Vascular Care

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
The measure specifications are very similar for three of the measure components, Daily Aspirin, Blood Pressure Control and Tobacco Free. However, the WCHQ measure also adds the Statin Use component which is a secondary prevention according to the AHA/ACC revised guidelines in November 2013. There are also some slight denominator differences in number and time frame of visits required.

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.)
Because this measure includes the secondary prevention element of Statin Use from the updated AHA/ACC guidelines from November 2013. It also uses a denominator algorithm that allows patient level lists to be generated for internal practice quality improvement purposes.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** RBS_File_Formats_060115.xlsx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Wisconsin Collaborative for Healthcare Quality
Co.2 **Point of Contact:** Mary, Gordon, mgordon@wchq.org, 608-775-4519-
Co.3 **Measure Developer if different from Measure Steward:** Wisconsin Collaborative for Healthcare Quality
Co.4 **Point of Contact:** Mary, Gordon, mgordon@wchq.org, 608-775-4519-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The WCHQ Measurement Advisory Committee (MAC). The MAC serves as a subcommittee of the Board of Directors for purposes of directing policy, actual changes to measures, and procedural or clinical decisions for WCHQ workgroups involved with new measures development or other measurement-related initiatives.

Member Names: Geoffrey Lamb, MD-Medical College of Wisconsin, Dan Collins -ThedaCare, Dirk Steinert, MD-Columbia St. Mary's, Greg Blommel, MD-Froedtert West Bend, Kristine Bruno, MD-Aurora Advanced, Robert Mead, MD-Bellin, Steve Kulick, MD-ProHealth Care, Kim Volberg – Dean, Rhonda Struckm MD-Wheaton Franciscan, John Zlabek, MD-Gundersen Health System.

Ambulatory Care Specifications (ACS) work group. The Ambulatory Care Specifications Workgroup develops, monitors, and revises the WCHQ ambulatory process and clinical outcome measure specifications. This work group is open to all WCHQ member, meets generally weekly and is generally attended by 18-20 member organizations and 30-40 people.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 10, 2014
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?** 10, 2015

Ad.6 **Copyright statement:** None
Ad.7 **Disclaimers:** Disclaimer: Measures reported by WCHQ healthcare organizations represent a specific aspect of care in relation to an evidence-based standard, but are not clinical guidelines and do not establish standards of care. All providers should have an individual care plan established with their patient.

Ad.8 **Additional Information/Comments:**

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 #: 2764
Measure Title: Fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure and LVEF <40% on ACEI or ARB and Beta-blocker Therapy
Measure Steward: National Minority Quality Froum
Brief Description of Measure: Percentage of patients aged 18 years and older with a diagnosis of heart failure (HF) and a current or prior ejection fraction (EF) <40% who are self-identified Black or African Americans and receiving ACEI or ARB and Beta-blocker therapy who were prescribed a fixed-dose combination of hydralazine and isosorbide dinitrate seen for an office visit in the measurement period in the outpatient setting or at each hospital discharge
Developer Rationale: The African-American Heart Failure Trial (A-HeFT) first published in 2004 demonstrated that there is significant benefit for African American patients who receive the fixed-dose combination therapy of hydralazine and isosorbide dinitrate. A-HeFT built on the findings from the two Vasodilator-Heart Failure Trials (V-HeFT). A-HeFT, which was ended early due to the mortality rates in the placebo population, demonstrated a 43% reduction in mortality, a 33% decrease in initial hospitalizations, and a 50% improvement in patient-reported quality of life (Taylor, 2004; Sharma, 2014). These results clearly demonstrate that the fixed-dose combination therapy significantly improves patient morbidity, mortality and quality of life in this clinical cohort. There is no substitute for the fixed-dose combination therapy.

Even with this strong evidence of unprecedented efficacy and cost-effectiveness, research shows that more than 85% of African American patients are not receiving the quality of care that this therapy affords, constituting a significant gap in care quality (Dickson, 2015). The underuse of the fixed-dose combination of hydralazine plus isosorbide dinitrate in African Americans with severe heart failure is a health care and health quality disparity that exposes these patients to an elevated risk for mortality and hospitalization, and compromises efforts to contain the escalating system costs by preventing or reducing unnecessary hospitalizations and readmissions.

Based upon research on the mortality benefit of the fixed-dose combination (Fonarow, 2011), the National Minority Quality Forum estimates that 51,542 (27%) of the 189,891 African American Medicare beneficiaries who were being treated for heart failure and received their prescription drugs under Part D should have been treated with the fixed-dose combination; but only 2,377 (5%) had at least one prescription (30-day supply) of the therapy. Further, the National Minority Quality Forum estimates that between 2008 and 2010, only 3% of the eligible patient cohort in Medicare received the therapy. Given the documented number to treat to receive the mortality benefit (21), it can be estimated that from 2007 through 2010, 20,000 African American Medicare beneficiaries died as a result of the failure to receive quality care as defined by evidence-based guidelines.

Research continues to explore if the fixed-dose combination of hydralazine and isosorbide dinitrate is linked to a particular genetic polymorphism (NIH funded Genomic Response Analysis of Heart Failure Therapy in African Americans). While we anticipate that the evidence supporting this treatment will be refined over time, the proven benefits to this patient population is significant and there is a clear opportunity for improvement. Failure to do so constitutes a failure to provide quality and cost-effective care.

References:

Dickson VV, Knafl GJ, Wald J, Riegel B. Racial differences in clinical treatment and self-care behaviors of adults with chronic heart failure. J Am Heart Assoc. 2015;4:1-13.
Fonarow GC, Yancy CW, Hernandez AF, Peterson ED, Spertus JA, Heidenreich PA. Potential impact of optimal implementation of evidence-based heart failure therapies on mortality. Am Heart J. 2011;161:1024-1030.

Sharma A, Colvin-Adams M, Yancy CW. Heart failure in African Americans: disparities can be overcome. Cleve Clin J Med. 2014;81:301-11.
Taylor AL, Ziesche S, Yancy C, et al. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. N Engl J Med 2004; 351:2049–57.

Numerator Statement: Patients prescribed a fixed-dose combination of hydralazine and isosorbide dinitrate seen for an office visit in the measurement period in the outpatient setting or at each hospital discharge
Denominator Statement: All patients aged 18 years and older with a diagnosis of heart failure with a current or prior EF <40% who are self-identified Black or African Americans and receiving ACEI or ARB and Beta-blocker therapy
Denominator Exclusions: Denominator exclusions include:

- Hypotension (severe or symptomatic)
- Severe lupus erythematosus
- Unstable angina
- Peripheral neuritis
- Patient actively taking Phosphodiesterase Type 5 (PDE5) Inhibitors

Measure Type: Process
Data Source: Electronic Clinical Data : Electronic Health Record
Level of Analysis: Clinician : Group/Practice, Clinician : Individual
Is this an eMeasure? ☒ **Yes** ☐ **No** **If Yes, was it re-specified from a previously endorsed measure?** ☐ **Yes** ☒ **No**

Is this a MAINTENANCE measure submission? ☐ **Yes** ☒ **No, this is a NEW measure submission.**
For a MAINTENANCE, what is the Original Endorsement Date: n/a **Most Recent Endorsement Date:** n/a

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measures evaluation and voting processes. The preliminary analysis will help to guide the Standing Committee evaluation of each measure by summarizing the measure developer submission, guide measure evaluation discussion, and identify topic areas for additional input. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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.

The developer provides the following evidence for this *process eMeasure*:

- This clinician-level process eMeasure calculates the percentage of patients aged 18 years and older with a diagnosis of heart failure (HF) and a current or prior ejection fraction (EF) <40% who are self-identified Black or African Americans and receiving ACEI or ARB and Beta-blocker therapy who were prescribed a fixed-dose combination of hydralazine and isosorbide dinitrate seen for an office visit in the measurement period in the outpatient setting or at each hospital discharge.
- The developer provides the 2013 ACCF/AHA guideline for the management of heart failure (**Class I; Level of Evidence: A**) with one recommendation for the combination of hydralazine and isosorbide dinitrate for patients self-described as African Americans (**Class I**) and the HFSA 2010 Comprehensive Heart Failure Practice Guideline (**Strength of Recommendation: Is Recommended**) with two recommendations: hydralazine and isosorbide dinitrate is recommended in addition to beta blockers and ACE inhibitors for African Americans with HF and reduced LVEF (**Strength of Evidence = A and B**) and hydralazine/isosorbide dinitrate is recommended for African American women with moderate to severe HF symptoms who are on background neurohormonal inhibition (**Strength of Evidence = B**).
- The evidence review supporting the hydralazine/isosorbide dinitrate recommendations was conducted through October 2011 and includes other references through April 2013 for the 2013 ACCF/AHA guideline. No information on the time period for the HFSA 2010 guideline was provided.

- QQC - 4 randomized controlled trials (RCTs) and 2 post hoc retrospective analyses supporting the 2013 ACCF/AHA guideline. No specific information on the number of studies included in the body of evidence for the HFSA 2010 Comprehensive Heart Failure Practice Guideline.
- Two additional analyses were published after the publication of the 2013 ACCF/AHA guideline and the developers conclude that, "While additional research on whether use of hydralazine and isosorbide dinitrate is linked to a genetic polymorphism may refine the clinical recommendations, findings in these publications further support the current recommendations and level of evidence ratings for the use of combination therapy in African American patients."
- The developer provides a diagram that demonstrates how the use of a fixed-dose combination of hydralazine and isosorbide dinitrate in self-identified black or African American patients with HF and LVSD receiving ACEI/ARB and beta-blocker therapy is linked to patient outcomes.

Questions for the Committee:

- *For process measures:*
  - *Is the evidence directly applicable to the process of care being measured?*
  - *Is the process of care proximal and closely related to desired outcomes?*
- *For possible exception to the evidence criteria:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empiric evidence?*

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

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

- Because this is a newly developed measure the developers do not yet have overall performance data or data on disparities from the measure as specified.
- The developer provides a summary of data from the literature that demonstrates the existence of a significant opportunity for improvement of whether eligible patients are receiving the hydralazine/isosorbide dinitrate combination therapy in the ambulatory setting and at hospital discharge.
- The developer provides a summary of data from the literature that demonstrates that HF is more prevalent in African Americans than in whites, occurs earlier, imposes higher rates of death and morbidity, and has a more malignant course. The developers also report that, "Much of the disparity can be assigned to modifiable risk factors such as uncontrolled hypertension and on suboptimal health care. Therefore, when African Americans are treated according to guidelines, discrepant outcomes can be minimized (Sharma, 2014)."

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*
- *Should this measure be indicated as disparities sensitive?*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b,)

1. ***Committee's Overview Comments***:
   - The evidence presented is from the 2013 ACCF/AHA clinical practice guidelines for managing heart failure as well as the Heart Failure Society of America's 2010 heart failure practice guidelines. ACCF/AHA recommendation is

Class I, Level A, and the HFSA recommendation is Strength of evidence level A for NYHA class 3 or 4, and level B for NYHA class 2. Evidence is based on 4 RCTs, one of which was conducted in AA, and found significant mortality benefit in this population with a 43% reduction in mortality and 33% relative reduction in hospitalizations. The evidence directly applies to the measure. The time period reviewed for the guidelines was through April 2013. Estimate of benefit (Fonarow) over 6000 lives saved per year, and a 50% improvement in QOL. QQC was provided from the 4 RCT's and 2 post-hoc analyses. An additional study was published that suggested a possible genetic link for use of this medication, and the authors felt that this further supports the recommendation. Suggests moderate evidence rating

1a. ***Committee's Comments on Evidence to Support Measure Focus:***

- The evidence directly relates to this process measure

1b. ***Committee's Comments on Performance Gap:***

- Performance gap was demonstrated to be quite large. A review of GWTG found only 7.3% of the AA HF population on this treatment compared to an estimated 27% that should have been. An analysis from the measure developer's estimate. The NNT is 21 for mortality benefit. Because this is a new measure submitted for trial request, the developers do not have specific data on performance gap identified by their e-measure.
- Disparities – this measure addresses a common condition that disproportionately affects a minority population and the treatment addressed by the measure has a significantly beneficial impact on the minority population. It addresses the National Quality Strategy of effective clinical care. Heart failure affects approximately 3% of the Black population and affects men and women approximately the same.

1c. ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications
2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- This individual and group clinician eMeasure assesses whether self identified Black or African American Heart Failure(HF) patients (identified by diagnosis, and LVEF < 40% or moderate or severe LVSD) who are currently taking ACEI/ARBs and beta-blockades are prescribed a fixed-dose combination of hydralazine and isosorbide dinitrate at an outpatient encounter or at hospital discharge.
- The measure's data source is an EHR. ICD 9, ICD 10, CPT, SNOMED, LOINC and prescription codes provided for the numerator and denominator for inpatient and outpatient setting. ICD-10 conversion methodology is not discussed. Higher scores equal better quality.
- The developers report that the measure is intended for use in a clinician office/clinic and hospital/acute care facility; however, the value sets include nursing facility, long-term residential facility and home care services which are not traditionally considered outpatient settings. If hospital discharges are included in the measure population, instructions for identifying the accountable clinician at the time of the hospital discharge is not provided.
- The HQMF specifications state that 2 or more outpatient (OP) encounters are required during the measurement year to establish a patient/provider relationship, rather than ≥ 2 encounters at any time.
- "Provider interactions" are listed as encounters in the value set spreadsheet and in the eMeasure specifications that include both face-to-face visits and non-face-to-face communications. The developer is encouraged to provide reasoning for inclusion, and clarify if all provider interactions are included in the denominator definition for a patient encounter.
- The HQMF denominator logic states that both ACEI/ARBs and beta-blockades should start before and overlap the HF encounter to be considered for a fixed-dose combination of hydralazine and isosorbide dinitrate therapy. The

developer is encouraged to clarify if both medications must be started before the 2nd encounter.

- Denominator exceptions include hypotension (severe or symptomatic), severe lupus erythematosus, unstable angina, peripheral neuritis, and patient actively taking Phosphodiesterase Type 5 (PDE5) Inhibitors. Hypotension is *not* also defined with diastolic and systolic parameters in the specifications.
- Missing numerator data present a quality failure.
- The calculation algorithm is included. The measure is not risk adjusted.
- The eMeasure specifications and values sets meet all current NQF eMeasure technical requirements and are provided on Sharepoint for SC review.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*
- *For an eMeasure:*
  - *Does the submission meet NQF's eMeasure criteria?*
  - *Are there questions or identified gaps with the value sets, data elements or HQMF logic?*
  - *Are there inconsistencies between the evidence, the Measure Information Form (MIF) and the eMeasure components or logic?*

2a2. Reliability Testing Testing attachment

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.

- In section3c.1, the developer indicates, "Because this measure is submitted for eMeasure trial approval and testing is not yet completed, we are not yet able to share information on data availability and collection beyond what is provided in the feasibility assessment in 3b.3. This information will be collected during our testing and modifications made to the measure based on the results."
- In addition to the above statement, the developer additionally submitted a reliability testing plan.
- The developer submitted pre-testing from the Measure Authoring Tool within the Bonnie Output that tests eMeasure logic. The measure logic successfully validated through the Bonnie Output.

Questions for the Committee:

- *Other specific questions regarding validity testing.*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The developer provides evidence that states Black or African American adults with HF who are currently taking ACEI/ARB and beta-blockade therapies should be also be taking a fixed-dose combination of hydralazine and isosorbide dinitrate.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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.

- In section3c.1, the developer indicates, "Because this measure is submitted for eMeasure trial approval and testing is not yet completed, we are not yet able to share information on data availability and collection beyond what is provided in the feasibility assessment in 3b.3. This information will be collected during our testing and modifications made to the measure based on the results."
- In addition to the above statement, the developer additionally submitted a validity testing plan.
- The developer submitted pre-testing from the Measure Authoring Tool within the Bonnie Output that also tests eMeasure performance calculation. This testing does use "live" EHR patients, though NQF currently accepts Bonnie Output pre-testing when EHR testing was not provided. Results in the 15 "pre-test" patients demonstrated 100% agreement for identifying both expected and actual initial patient population, denominator, denominator exclusions, numerator, and denominator exceptions. Testing characteristics are provided for the 15 "pre-test" patients, with 76% of all possible data elements included in the pre-test sample.

Questions for the Committee:

- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Denominator exceptions include hypotension (severe or symptomatic), severe lupus erythematosus, unstable angina, peripheral neuritis, and patient actively taking Phosphodiesterase Type 5 (PDE5) Inhibitors.
- Impact of exclusions on the measure to be provided when testing is completed.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment:

- This process measure is not risk adjusted.
- The developer did not consider either clinical or SDS adjustment for this measure.

2b5. Meaningful difference:

- A testing plan was provided.

2b6. Comparability of data sources/methods:

- This section is not applicable as the developer submitted a single specification (eMeasure) with one data source (EHR).

2b7. Missing Data

- Missing numerator data present a quality failure.
- A testing plan was provided.

Committee pre-evaluation comments

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

2a1. &2b1.: ***Committee's Comments on Reliability-Specifications:***

- All data elements are defined with VSAC registered value sets, and is specified in the HQMF format using the QDM. Again, because it is submitted for trial use testing is not complete. They did submit a testing plan that appears to comply with all required testing. The measure is intended for the outpatient population and the hospital acute care population. Bonnie output from pre-testing successfully validated the measure logic, with a 100% pass rate, for 85% of the data elements that were covered in the testing.
- The exclusions are identical to that listed in the FDA prescribing information as contraindications and/or warnings. Severe or symptomatic heart failure was defined using codes – not by a defined BP level, which seems appropriate. The only concern identified was the potential for missing information on ejection fraction in the EHR. One question for the developer is whether there is any concern in interpreting adherence to the measure based on what has been noted in the literature of poor tolerance/adherence to the medication. The measure logic requires the patient to have at least 2 encounters during the measurement year to establish a relationship. Those encounters can be in any setting including electronic/phone encounters. Specifications are

consistent with the measure description. Reliability testing is not required for this measure as it was submitted for trial approval. Reliability testing plan included and will address all data elements and the effect of missing data, and compare to chart abstracted records, using Cohen's Kappa.

2a2.: ***Committee's Comments on Reliability-Testing:***

- Reliability testing plan included and will address all data elements and the effect of missing data, and compare to chart abstracted records, using Cohen's Kappa. Not required for a trial measures

2b1.: ***Committee's Comments on Validity-Specifications***:

- It was tested in a test EHR population size 15 and had 76% of data elements covered with a 100% pass rate. The testing plan specifies the number needed to test to assure 80% power to test for differences in Kappa statistics and performance rates. Measure testing will compare chart abstraction vs. electronic abstraction, sensitivity and specificity. They will also test for the frequency of exclusions to address threats to validity in more than one EHR. The measure is not risk-adjusted. Testing not required for a trial measure.

2b2.: ***Committee's Comments on Validity-Testing:***

- It was tested in a test EHR population size 15 and had 76% of data elements covered with a 100% pass rate. The testing plan specifies the number needed to test to assure 80% power to test for differences in Kappa statistics and performance rates. Measure testing will compare chart abstraction vs. electronic abstraction, sensitivity and specificity. They will also test for the frequency of exclusions to address threats to validity in more than one EHR. The measure is not risk-adjusted. Testing not required for a trial measure.

2b3-7.: ***Committee's Comments on Threats to Validity:***

- Potential threats to validity would be missing data and possibly the fact that patient refusal or allergies and other patients reasons are not exclusions. It was noted that many people are unable to comply with the dosing regimen.
- The measure is not risk adjusted - it is intended for a minority population.

2d.: ***Committee's Comments on Composite Performance Measure:***

- Not Applicable

Criterion 3. Feasibility

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

- The eMeasure is specified for use with an EHR as the data source.
- The developers provide an eMeasure Feasibility Scorecard of 2 EHRS (hospital and outpatient) testing all data elements required to calculation the measure. All data elements for both EHRs scored 3s (except Ejection Fraction < 40%) meaning the data elements are routinely collected as part of routine care and require no additional data entry from the clinician for the quality measure and no EHR user interface changes. Ejection Fraction <40% scored 2 in data standards meaning data element is not routinely collected as part of routine care and additional time and effort over and above routine care is required, but perceived to have some benefit.
- Per the Writing Committee recommendation, the developer changed "Diagnosis of Worsening Ischemic Heart Disease* to "Unstable Angina" better identify applicable patients for fixed combination therapy. As unstable angina is seen as a subset of ischemic heart disease, the feasibility rating would not be impacted.
- The developer provides the measure specifications free of charge to provider end users.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems*

and sites?

Committee pre-evaluation comments
Criteria 3: Feasibility

3.: ***Committee's Comments on Feasibility:***

- Developers provide a feasibility scorecard from 1 inpatient and 1 outpatient EHR from the same vendor. Scorecard measured data availability, data accuracy, data standards, workflow and a total data element feasibility score for all data elements. Each data element scored at the highest level with the exception of ejection fraction < 40%, which scored 11 out of 12. Testing will continue. This meets the requirements for a trial measure.

Criterion 4: Usability and Use

4. Usability and 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.

- The developer provides plans for future accountability and quality improvement use for this new eMeasure.
- As the measure is newly developed, the developer states unintended consequences have yet to be identified.

Questions for the Committee *:*

- *Is the measure publicly reported?*
- *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?*

Committee pre-evaluation comments
Criteria 4: Usability and Use

4.: ***Committee's Comments on Usability and Use:***

- As a new measure, it is intended for use at the clinician level. If successful in the trial, they plan to submit for CMS use under PQRS, possibly as early as 2017. A similar measure is currently used in the AHA's GWTG program for QI and benchmarking. A publication from that data suggested that 22% of eligible patients discharged from the hospital were discharged on the medication. This meets the requirements for a trial measure.

Criterion 5: Related and Competing Measures

Currently endorsed measures:

- 0081 : Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)
- 0083 : Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD)

Previously endorsed measures:

- 0162: ACEI or ARB for left ventricular systolic dysfunction - Heart Failure (HF) Patients (CMS)
- 0610: Heart Failure - Use of ACE Inhibitor (ACEI) or Angiotensin Receptor Blocker (ARB) Therapy (ActiveHealth Management)
- 0615: Heart Failure - Use of Beta Blocker Therapy (ActiveHealth Management)

- The developer reports that measure specifications for the target population and medication therapies for ACEI, ARB, and beta-blocker are completely harmonized with 0081 and 0083.

Pre-meeting public and member comments

Comment by: David Smith
Organization: Public – Yale University
Comment#5147: I am writing in support of a proposed quality measure that has the potential to save thousands of lives annually by highlighting a preventable treatment deficiency, namely, the National Minority Quality Forum's submission (# 2764) regarding a fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure (HF) and LVEF <40% on ACEI or ARB and Beta-blocker Therapy.
Today, less than 10% of eligible heart failure patients are being prescribed an FDA-approved treatment that's been proven to significantly reduce hospitalization and mortality rates. That's why I'm writing in support of the measure submitted by the National Minority Quality Forum (NMQF) that strongly encourages healthcare providers to ensure that eligible African American patients with heart disease receive the proper course of treatment. More despairingly is the fact that our current trainees are learning little about this treatment opportunity in their current curricula. As a professor, it is most alarming that other teachers and attending professionals do not know how to adequately prescribe or dose the medicines appropriately and that there IS NO GENERIC EQIVALENT. So, the perpetuation of this type of neglect has vast repercussions and dreadful prediction for the future that immediate address of this problem promises immense future returns.
The science behind the impact of this FDA-approved drug has been well documented. Its benefits have been published in the New England Journal of Medicine and other peer-reviewed sources, and the American College of Cardiology and American Heart Association have released detailed practice guidelines calling for this specific treatment protocol. Nonetheless, while published studies estimate that there are over 150,000 African Americans living in America who could benefit from this treatment, only 7% (or 11,000) of them are receiving it. As a consequence, experts have estimated that 6,655 blacks die prematurely every year.
An endorsement from the National Quality Forum (NQF) is considered the highest standard for healthcare quality, and sends a strong message to providers that measures are evidence-based, valid, and can help patients achieve better outcomes. I strongly believe that the proposed heart failure measure meets NQF's criteria, and encourage you to provide your formal endorsement in order to help facilitate widespread adoption of this treatment.
I appreciate the opportunity to weigh in on this important issue, and urge NQF to approve this quality measure submission. If I may lend any further words of support, please do not hesitate to call me.

David N Smith, MD
Clinical Assistant Professor, Yale University

Comment by: Mr. Joseph Harris
Organization: Public
Comment#5146: I am writing in support of proposed quality measure 2764 that has the potential to save thousands of lives.

Comment by: Anekwe Onwuanyi, MD
Organization: Public
Comment#5143: I am writing in support of a proposed quality measure that has the potential to save thousands of lives annually by highlighting a preventable treatment deficiency, namely, the National Minority Quality Forum's submission (#2764) regarding a fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure (HF) and LVEF <40% on ACEI or ARB and Beta-blocker Therapy.
Today, less than 10% of eligible heart failure patients are being prescribed an FDA-approved treatment that's been proven to significantly reduce hospitalization and mortality rates. That's why I'm writing in support of the measure submitted by the National Minority Quality Forum (NMQF) that strongly encourages healthcare providers to ensure that eligible African American patients with heart disease receive the proper course of treatment.
The science behind the impact of this FDA-approved drug has been well documented. Its benefits have been published in the New England Journal of Medicine and other peer-reviewed sources, and the American College of Cardiology and

American Heart Association have released detailed practice guidelines calling for this specific treatment protocol. Nonetheless, while published studies estimate that there are over 150,000 African Americans living in America who could benefit from this treatment, only 7% (or 11,000) of them are receiving it. As a consequence, experts have estimated that 6,655 blacks die prematurely every year.
An endorsement from the National Quality Forum (NQF) is considered the highest standard for healthcare quality, and sends a strong message to providers that measures are evidence-based, valid, and can help patients achieve better outcomes. I strongly believe that the proposed heart failure measure meets NQF's criteria, and encourage you to provide your formal endorsement in order to help facilitate widespread adoption of this treatment. I appreciate the opportunity to weigh in on this important issue, and urge NQF to approve this quality measure submission.

Comment by: Mr. Adolph P. Falcon, MPP
Organization: Public
Comment#5142: American patients with heart disease get access to the right drug for them. For this reason, the National Alliance for Hispanic Health offers its full support for proposed quality measure #2764.

Comment by: Mr. Joseph Earl Harris, Jr.
Organization: Public
Comment#5141: I am writing in support of a proposed quality measure that will save thousands of lives. Although clinical trial evidence supports the use of fixed dose hydralazine and isosorbide dinitrate to improve survival in African Americans with advanced heart failure, less than 10 percent of eligible patients receive this therapy. I encourage NQF to provide its formal endorsement in order to facilitate widespread adoption of this treatment.

Comment by: Michele Blair
Organization: Public
Comment#5138: "We support the combined use of hydralazine and isosorbide dinitrate for self-identified Black or African American patients with heart failure (HF) and reduced ejection fraction on ACE inhibitor and beta-blocker therapy. As stated in our national guideline: A combination of hydralazine and isosorbide dinitrate is recommended as part of standard therapy in addition to beta blockers and ACE inhibitors for African Americans with LV systolic dysfunction and: - New York Heart Association (NYHA) class III or IV HF (Strength of Evidence = A)

- NYHA class II HF (Strength of Evidence = B)

Comment by: Modele O. Ogunniyi
Organization: Public
Comment#5136: An endorsement from the National Quality Forum (NQF) is considered the highest standard for healthcare quality, and sends a strong message to providers that measures are evidence-based, valid, and can help patients achieve better outcomes. I strongly believe that the proposed heart failure measure meets NQF's criteria, and encourage you to provide your formal endorsement in order to help facilitate widespread adoption of this treatment. Thanks for the opportunity to comment on this important issue, and urge NQF to approve this quality measure submission.

Comment by: Beverly E. Oliver, FNP-BC
Organization: Public
Comment#5135: I am writing in support of a proposed quality measure that has the potential to save thousands of lives annually by highlighting a preventable treatment deficiency, namely, the National Minority Quality Forum's submission (# 2764) regarding a fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure (HF) and LVEF <40% on ACEI or ARB and Beta-blocker Therapy.
Today, less than 10% of eligible heart failure patients are being prescribed an FDA-approved treatment that's been proven to significantly reduce hospitalization and mortality rates. That's why I'm writing in support of the measure submitted by the National Minority Quality Forum (NMQF) that strongly encourages healthcare providers to ensure that eligible African American patients with heart disease receive the proper course of treatment.
The science behind the impact of this FDA-approved drug has been well documented. Its benefits have been published

in the New England Journal of Medicine and other peer-reviewed sources, and the American College of Cardiology and American Heart Association have released detailed practice guidelines calling for this specific treatment protocol. Nonetheless, while published studies estimate that there are over 150,000 African Americans living in America who could benefit from this treatment, only 7% (or 11,000) of them are receiving it. As a consequence, experts have estimated that 6,655 blacks die prematurely every year.
An endorsement from the National Quality Forum (NQF) is considered the highest standard for healthcare quality, and sends a strong message to providers that measures are evidence-based, valid, and can help patients achieve better outcomes. I strongly believe that the proposed heart failure measure meets NQF's criteria, and encourage you to provide your formal endorsement in order to help facilitate widespread adoption of this treatment.
I appreciate the opportunity to weigh in on this important issue, and urge NQF to approve this quality measure submission.
Beverly Oliver, FNP-BC

Comment by: Jan Neil Basile, MD
Organization: Public - Medical University of South Carolina
Comment#5130: I am a clinical trialist and hypertension specialist who cares for patients with heart failure including a large panel of AA patients.
Health care disparities continue to exist in minority populations for many reasons including clinical access, formulary availability, mistrust, socioeconomic position, and cultural and language issues just to mention a few. When there is solid clinical trial evidence of outcome benefit in a minority population given a IA recommendation of benefit from the American Heart Association (2014), quality measures should ensure that clincians are held to this standard of care when having the opportunity to treat such patients.
Based on the African American Heart Failure Trial (AHEFT) in 2004 published in the NEJM, self described blacks who had heart failure (HF) with a reduced ejection fraction gained a significant 43% reduction in death, a reduction in first or recurrent hospitalization for HF, and an improvement in quality of life when fixed-dose isosorbide dinitrate/hydralazine was added to an ACE inhibitor or ARB + a Beta Blocker. This is believed to occur because of a unique pathophysiologic derangement in nitric oxide upregulation in Blacks.
By making isosorbide dinitrate/hydralaziine a quality improvement measure in blacks who meet the definition of the AHEFT clinical trial we will ensure that this minority population who face tremendous obstacles from the social determinants of health to at least be assured of getting the best evidence base for clinical care.
Jan Basile, MD
Professor of Medicine
Medical University of South Carolina
Charleston, SC

Comment by: Dr. David A. Mann, MD, PhD
Organization: Public
Comment#5129: Representing my own opinions, and not that of any organizations that I work for or are affiliated with, I support the aim of proposed measure 2764 but not necessarily its exact language.
I fully support the goal of providing Black heart failure patients with optimal therapy for their heart failure. And I think that combination therapy with hydralazine and isosorbide dinitrate in the setting described in the measure has good evidence behind it.
I am not sure, however, that medical science is certain that only the specific dose combination used in the A-HeFT trial, provided in one particular proprietary combination formulation, is effective for this indication. Therefore I am hesitant to endorse a measure that appears to require the use of one particular proprietary formulation.
Immediate prescription of this fixed dose proprietary product bypasses dose titration, does not allow for individualized therapy, precludes use of more affordable generics, and may potentially generate more adverse effects than would occur with individualized dose titration to this therapeutic goal. I don't think that represents optimal care for patients.
How does the current language match up with underlying intent? If the intent is to encourage combination therapy with these two agents without requiring a particular product or a particular dosage, the language seems too restrictive. If the intent is to encourage the exact doses used in the A-HeFT trial, then the language is too lenient: any fixed dose

combination at any doses of the agents would meet the stated measure.
As a quality measure, the current language could be problematic. If a patient is titrated to 100% of the A-HeFT dose of agent 1 but only tolerates 75% of the dose of agent 2, is that patient a fail on this metric? I would hope not, but by its exact language, the answer would seem to be yes.

Comment by: Elizabeth O. Ofili, MD
Organization: Public
Comment#5123: I am a cardiologist in clinical practice with a large number of African American patients. I see the daily struggles of patients whose quality of life is deeply impacted. The African American Heart Failure Trial (AHEFT) was prematurely stopped by the DSMB and published in NEJM in 2004. This landmark study showed that self-described African Americans or Blacks, had over 40% survival, as well as hospitalization and quality of life benefits when treated with fixed dose combination of isosorbide dinitrate and hydralazine (FDC I/H) on top of standard therapy.The evidence was so strong that it received a level 1A by the guideline writing committee and has been affirmed by each committee since then. It is a health equity issue that the most recent analysis of America's superior hospitals, show that very few African American patients are receiving this therapy. I join with others concerned with health disparities and the attainment of health equity, in asking NQF to add FDC I/H as a standard of care for self-described African Americans, as contained in every heart failure guideline since 2004.Thank you for helping us to deliver quality heart failure care for our patients. Elizabeth Ofili, MD, MPH, FACC Professor of Medicine and Attending Cardiologist

Comment by: Dr. Traci Ferguson, WellCare; Ms. Kiersten Adams
Organization: HPL- WellCare Health Plans
Comment#5121: WellCare Health Plans, Inc. fully supports the endorsement of NQF quality measure #2764, "Fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure and LVEF <40% on ACEI or ARB and Beta-blocker Therapy." The benefits of combining Hydralazine and Isosorbide Dinitrate have been published in various peer-reviewed sources, including the New England Journal of Medicine. Additionally, this approach is supported by both the American College of Cardiology and the American Heart Association.
As one of the country's largest health care companies dedicated solely to serving public program beneficiaries, we see the effects that disparities can have on health outcomes. Adoption of this measure will ensure that eligible African American patients with symptomatic heart failure receive the proposed course of treatment. WellCare believes that endorsement of this quality measure submitted by the National Minority Quality Forum will increase the utilization of this evidence-based standard of care, thus saving thousands of lives each year.

Comment by: Tamarah Duperval-Brownlee, MD, MPH, MBA
Organization: Public
Comment#5120: An endorsement from the National Quality Forum (NQF) is considered the highest standard for healthcare quality, and sends a strong message to providers that measures are evidence-based, valid, and can help patients achieve better outcomes. I strongly believe that the proposed heart failure measure meets NQF's criteria, and encourage you to provide your formal endorsement in order to help facilitate widespread adoption of this treatment.
I appreciate the opportunity to weigh in on this important issue, and urge NQF to approve this quality measure submission.

Comment by: Cassandra McCullough; Mr. Andrew M. Rosenberg
Organization: Association of Black Cardiologists
Comment#5117: On behalf of over 1,500 healthcare professionals dedicated to treating patients with cardiovascular disease and to achieving health equity for all through the elimination of disparities, we are writing to express our strong support of quality measure #2764 to promote the most effective course of treatment for eligible African Americans with heart failure (HF).
Founded in 1974, the ABC is a nonprofit organization with an international membership comprised of health professionals, lay members of the community (Community Health Advocates), corporate members, and institutional members. At the Association of Black Cardiologists (ABC), there is no issue more central to our cause than ensuring that all Americans are given the foremost care to combat, treat, and overcome cardiovascular disease. This includes the

recognition that cardiovascular disease occurs disproportionately in African Americans. The National Minority Quality Forum's (NMQF's) recently proposed quality measure represents a critical step towards furthering these goals, and we hope that you will join us in encouraging its widespread adoption by providers across the country.
The ABC is not new to this issue, indeed, our organization played a key role in the execution of the landmark African-American Heart Failure Trial (A-HeFT) that provided the clinical evidence upon which the NMQF's proposed quality measure is based. That data was published in 2004 in the New England Journal of Medicine as "breaking news," and was highlighted later that year at the annual American Heart Association Scientific Meeting.
We recall that the A-HeFT trial was terminated prematurely due to the significant outcomes present in the treated group. In fact, the results were so positive, the FDA's Data Safety Monitoring board deemed it unethical to allow the untreated group to proceed without the opportunity to receive this profound benefit.
The merit of this proposed measure—and our support of it—is defined by hard data and indisputable evidence: the A-HeFT study demonstrated that its fixed-dose standard of care reduced mortality rates in African Americans with heart failure by over 40% while also significantly reducing first-time hospitalizations. Yet despite the consensus that emerged from the medical community on the regimen's benefits, today it reaches only 7% of over 150,000 clinically-eligible African American patients across the country.
This concern should not be unique to ABC, NMQF, and other organizations focused on eliminating healthcare disparities. Instead, this issue speaks to anyone who believes that standards of care should be evidence-based, valid, and help patients achieve better outcomes. The role of our organization is to advocate for the cardiovascular treatments that will help all patients live fuller and longer lives, but nowhere is this more important than in our efforts to address disparities among people of color.

Comment by: Mr. Ilen Bell
Organization: Public
Comment#5112: As a co-founder of Black Fitness Today, a leader in promoting health in the African American community, I am writing in support of a proposed quality measure that has the potential to save thousands of lives annually by highlighting a preventable treatment deficiency, namely, the National Minority Quality Forum's submission (# 2764) regarding a fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure (HF) and LVEF <40% on ACEI or ARB and Beta-blocker Therapy.. When considering the number of lives that can potentially be saved annually -- over the past decade, approximately 66,550 African Americans have perished without being provided the opportunity to choose Hydralazine and Isosorbide Dinitrate Therapy.
It is alarming that only "10% of eligible heart failure patients are being prescribed this FDA-approved treatment," which "has been proven to reduce mortality in blacks by 43% and first-time hospitalizations for HF by 38%." That's why I'm writing in support of the measure submitted by the National Minority Quality Forum (NMQF) that strongly encourages healthcare providers to ensure that eligible African American patients with heart disease receive the proper course of treatment.

Comment by: Oladipupo Olafiranye
Organization: Public
Comment#5111: "I am writing in strong support of the National Minority Quality Forum's (NMQF's) submission regarding a fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure and left ventricular ejection fraction of less 40% on ACEI or ARB and Beta-blocker Therapy. As a member of the Association of Black Cardiologists (ABC), I strongly believe that this quality measure has the potential to save thousands of lives annually by highlighting a preventable treatment deficiency.Although, heart failure affects millions of Americans, African American are disproportionately affected by heart failure with age-adjusted death rates remaining higher in African Americans than other populations. And despite the fact that there is an FDA-approved treatment that has been proven to be particularly effective in African Americans, only a small percentage of those who are clinically eligible are receiving the treatment."

Comment by: David Maron, MD
Organization: Public
Comment#5107: "Although clinical trial evidence supports the use of fixed-dose hydralazine and isosorbide dinitrate to

improve survival in African Americans with advanced heart failure, less than 10% of eligible patients receive this therapy. This proposed quality measure will raise awareness and increase the appropriate treatment of eligible African American patients with heart failure."

Comment by: James Januzzi, Jr., MD
Organization: Public - Harvard Medical School
Comment#5094: "As a clinician and clinical trialist, I am amazed at the gap between trial results and real-world prescription of a life-saving therapy for patients with HF such as we see with the under-use of hydralazine/nitrates in Blacks/African Americans. I agree this is a hugely important topic in need of further scrutiny and comment.
James Januzzi, MD; Professor of Medicine, Harvard Medical School."

Comment by: LaVarne Burton; Mr. Michael Spigler
Organization: American Kidney Fund
Comment#5049: "The American Kidney Fund (AKF) offers its full support of NQF# 2764. AKF is dedicated to ensuring that every kidney patient has access to health care, and that every person at risk for kidney disease is empowered to prevent it. As the nation's largest not-for-profit organization serving people with, and at risk for, kidney disease, we have helped more than 1 million low-income dialysis patients to access lifesaving medical care since our founding in 1971.
There are currently 31 million Americans living with some level of chronic kidney disease (CKD). Of these 31 million, minority populations face a greater risk of progressing from early CKD to kidney failure. African Americans with CKD, in particular, are disproportionately affected. More than 1 in 3 kidney failure patients living in the United States is African American.
Several studies have also shown that heart disease is a primary risk factor for developing kidney failure. That means that for the estimated 150,000 African Americans living with heart failure (HF), their risk for ultimately developing kidney failure is even greater.
AKF is committed to eliminating health disparities in CKD. The benefits of fixed-dose hyralazine and isosorbide dinitrate have been published in the New England Journal of Medicine and other peer-reviewed sources, yet only 7% of clinically eligible African Americans receive the treatment. We believe that adoption of this quality measure will improve African Americans' access to this life-saving treatment and will not only directly improve the outcomes for HF, but also indirectly improve the outcomes for African Americans at-risk for CKD."

Comment by: Nilam Sheth, PharmD
Organization: Public
Comment#5039: "I am writing in support of a proposed quality measure that has the potential to save thousands of lives annually by highlighting a preventable treatment deficiency, namely, the National Minority Quality Forum's submission (# 2764) regarding a fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure (HF) and LVEF <40% on ACEI or ARB and Beta-blocker Therapy.
Today, less than 10% of eligible heart failure patients are being prescribed an FDA-approved treatment that's been proven to significantly reduce hospitalization and mortality rates. That's why I'm writing in support of the measure submitted by the National Minority Quality Forum (NMQF) that strongly encourages healthcare providers to ensure that eligible African American patients with heart disease receive the proper course of treatment.
The science behind the impact of this FDA-approved drug has been well documented. Its benefits have been published in the New England Journal of Medicine and other peer-reviewed sources, and the American College of Cardiology and American Heart Association have released detailed practice guidelines calling for this specific treatment protocol.
Nonetheless, while published studies estimate that there are over 150,000 African Americans living in America who could benefit from this treatment, only 7% (or 11,000) of them are receiving it. As a consequence, experts have estimated that 6,655 blacks die prematurely every year.
An endorsement from the National Quality Forum (NQF) is considered the highest standard for healthcare quality, and sends a strong message to providers that measures are evidence-based, valid, and can help patients achieve better outcomes. I strongly believe that the proposed heart failure measure meets NQF's criteria, and encourage you to provide your formal endorsement in order to help facilitate widespread adoption of this treatment.
I appreciate the opportunity to weigh in on this important issue, and urge NQF to approve this quality measure submission."

Comment by: Chris Adamec, MPA
Organization: Public -The Healthcare Leadership Council (HLC)
Comment#5038: "Comments on Fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure and LVEF <40% on ACEI or ARB and Beta-blocker Therapy
The Healthcare Leadership Council (HLC) respectfully submits these comments in connection with NQF 2015 Cardiovascular Project. In this response, we support efforts by the National Minority Quality Forum (NMQF), the Association of Black Cardiologists (ABC), and other stakeholders to support the proposed measure that would support fixed-dose hyralazine and isosorbide dinitrate for self - identified Black or African American patients with heart failure. As you may be aware, today, only a very small number (about 7%) of African Americans who are clinically eligible for the FDA-approved therapy are getting it. As a consequence, over 6,500 blacks die prematurely every year because they are not receiving or adhering to standard of care. The quality measure would act to strongly encourage healthcare providers to ensure that eligible African American patients with heart disease receive the proper course of care treatment.
HLC, a coalition of chief executives from all disciplines within American healthcare, is the exclusive forum for the nation's healthcare leaders to jointly develop policies, plans, and programs to achieve their vision of a 21st century system that makes affordable, high-quality care accessible to all Americans. Members of HLC – hospitals, academic health centers, health plans, pharmaceutical companies, medical device manufacturers, biotech firms, health product distributors, pharmacies, and information technology companies – envision a quality-driven system that fosters innovation. HLC members advocate measures to increase the quality and efficiency of American healthcare by emphasizing wellness and prevention, care coordination, and the use of evidence-based medicine, while utilizing consumer choice and competition to elevate value.
We encourage NQF to endorse quality measure #2764, "Fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure and LVEF <40% on ACEI or ARB and Beta-blocker Therapy." The benefits of this approach have been published in the New England Journal of Medicine and other peer-reviewed sources. They also align with guidelines from the American College of Cardiology and the American Heart Association.
HLC appreciates this opportunity to comment on the proposed Cardiovascular measures. We believe there is tremendous potential for the health care industry as a whole to bring about robust collaboration and quality improvement in achieving our shared goal of improving the value of healthcare delivery for all.**"**

Comment by: Mr. Juan M. Cofield
Organization: NAACP Board of Directors
Comment#5037: "The NAACP Board of Directors adopted Health equality for all Americans includig a healthy life and high-quality health care as one of 5 Game changers. In support of the this Game Changer, the New England Area Conference (NEAC) of the NAACP strongly supports and advocates that African Americans who are clinical eligible for the FDA-approved therapy for Heart Failure. Further, NEAC encourages healthcare providersensure that eligible African American patients with heart disease receive the proper course of care treatment - namely, the fixed dose of Hydralazine and isosobide dinitrate."

Comment by: Mark S. Johnson
Organization: Howard University
Comment#5154: I would like to comment about the above referenced recommendation. While I agree that it would be useful to have more African American HF patients take Hydralazine and Isosorbide Nitrate as part of the HF arsenal, I strongly reject the recommendation that this only be given in the fixed dose combination that is currently on the market. In my clinical experience few patients, especially elderly patients have been able to tolerate the fixed combination dose. The mean age of the patients who were in the NEJM article was the 57. Even in these patients the side effects rates were high (48% had headache and 27% had dizziness). Only 68% were able to reach target dose.
It is possible that the current fixed dose was chosen to avoid generic duplication. Patients should be given lower doses and titrated slowly.
Mark Johnson, MD MPH
Professor, Community and Family Medicine

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Heart Failure: Fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black and African American Patients with Heart Failure and LVEF <40% on ACEI or ARB and Beta-blocker Therapy

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

Date of Submission: Click here to enter a date

Instructions

- *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.*
- Respond to all questions as instructed with answers immediately following the question. 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.
- Maximum of 10 pages (*incudes 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.

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

Evidence to Support the Measure Focus

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

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- 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.

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 U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Fixed-dose combination therapy of hydralazine and isosorbide dinitrate therapy for self-identified Black or African American patients with HF, LVSD and on ACEI or ARB and beta-blocker therapy

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

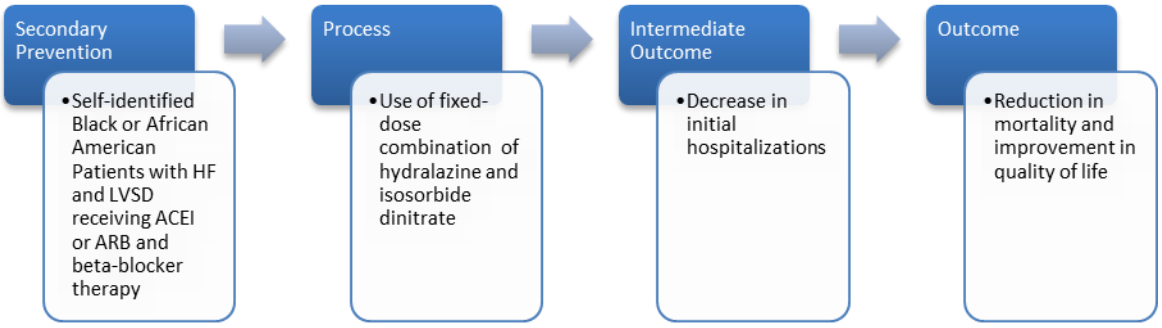
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



The African-American Heart Failure Trial (A-HeFT) first published in 2004 demonstrated that there is significant benefit for African American patients who receive a fixed-dose combination therapy of hydralazine and isosorbide dinitrate. This trial built on the findings from the two Vasodilator-Heart Failure Trials (V-HeFT). A-HeFT, which was ended early due to the mortality rates in the placebo population, demonstrated a 43% reduction in mortality, 33% decrease in initial hospitalizations, and a 50% improvement in patient-reported quality of life (Taylor, 2004, Sharma, 2014). These results clearly demonstrate that the fixed-dose combination therapy significantly improves patient morbidity, mortality and quality of life in this clinical cohort. There is no substitute for the fixed-dose combination therapy.

References:

Sharma A, Colvin-Adams M, Yancy CW. Heart failure in African Americans: disparities can be overcome. Cleve Clin J Med. 2014;81:301-11.

Taylor AL, Ziesche S, Yancy C, et al. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. N Engl J Med 2004; 351:2049–57.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections*** *1a.4* ***, and*** *1a.7*

☐ US Preventive Services Task Force Recommendation – ***complete sections*** *1a.5* ***and*** *1a.7*

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections*** *1a.6* ***and*** *1a.7*

☐ Other – ***complete section*** *1a.8*

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

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.

http://content.onlinejacc.org/article.aspx?articleid=1695825

Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, Givertz MM, Klapholz M, Moser DK, Rogers JG, Starling RC, Stevenson WG, Tang WHW, Teerlink JR, Walsh MN. Executive Summary: HFSA 2010 Comprehensive Heart Failure Practice Guideline. J Card Fail 2010;16:475-539. http://www.hfsa.org/hfsa-wp/wp/heart-failure-guidelines-2/.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

2013 ACCF/AHA Guideline for the Management of Heart Failure (e179)

The combination of hydralazine and isosorbide dinitrate is recommended to reduce morbidity and mortality for patients self-described as African Americans with NYHA class III–IV HFrEF receiving optimal therapy with ACE inhibitors and beta blockers, unless contraindicated. (**Class I; Level of Evidence: A**)

HFSA 2010 Comprehensive Heart Failure Practice Guideline

p. e80-81:

A combination of hydralazine and isosorbide dinitrate is recommended as part of standard therapy in addition to beta blockers and ACE inhibitors for African Americans with HF and reduced LVEF.

- NYHA III or IV HF (**Strength of Evidence = A**)
- NYHA II HF (**Strength of Evidence = B**)

p. e171:

The combination of hydralazine/isosorbide dinitrate is recommended as standard therapy for African American women with moderate to severe HF symptoms who are on background neurohormonal inhibition. (**Strength of Evidence = B**)

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

2013 ACCF/AHA Guideline for the Management of Heart Failure

Class of Recommendation: Class I

Definitions:

Class of Recommendation (COR) is an estimate of the size of the treatment effect considering risks versus benefits in addition to evidence and/or agreement that a given treatment or procedure is or is not useful/effective or in some situations may cause harm.

Class I: Procedure/Treatment should be performed/administered

HFSA 2010 Comprehensive Heart Failure Practice Guideline

Strength of Recommendation: Is recommended

Definition: The phrase ‘‘is recommended’’ should be taken to mean that the recommended therapy or management process should be followed as often as possible in individual patients. Exceptions are carefully delineated.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

Class of Recommendation (COR) is an estimate of the size of the treatment effect considering risks versus benefits in addition to evidence and/or agreement that a given treatment or procedure is or is not useful/effective or in some situations may cause harm.

Class I: Procedure/Treatment should be performed/administered

Class IIa: It is reasonable to perform procedure/administer treatment

Class IIb: Procedure/Treatment may be considered

Class III: No benefit (Not helpful or No proven benefit)

Class III: Harm (Excess cost w/o benefit or Harmful to patients)

HFSA 2010 Comprehensive Heart Failure Practice Guideline

Strength of Recommendation:

The HFSA guideline employs the categorization for strength of recommendation outlined in Table 1.3. There are several degrees of favorable recommendations and a single category for therapies felt to be not effective. The phrase ‘‘is recommended’’ should be taken to mean that the recommended therapy or management process should be followed as often as possible in individual patients. Exceptions are carefully delineated. ‘‘Should be considered’’ means that a majority of patients should receive the intervention, with some discretion involving individual patients. ‘‘May be considered’’ means that individualization of therapy is indicated (Table 1.3). When the available evidence is considered to be insufficient or too premature, or consensus fails, issues are labeled un-resolved and included as appropriate at the end of the relevant section.

Is recommended: Part of routine care; exceptions to therapy should be minimized

Should be considered: Majority of patients should receive the intervention; some discretion in application to individual patients should be allowed

May be considered: Individualization of therapy is indicated

Is not recommended: Therapeutic intervention should not be used

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

ACCF/AHA Task Force on Practice Guidelines. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines. American College of Cardiology Foundation and American Heart Association, Inc. Cardiosource.com. 2010. Available at:

http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf and

http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, Givertz MM, Klapholz M, Moser DK, Rogers JG, Starling RC, Stevenson WG, Tang WHW, Teerlink JR, Walsh MN. Executive Summary: HFSA 2010 Comprehensive Heart Failure Practice Guideline. J Card Fail 2010;16:475-539. http://www.hfsa.org/hfsa-wp/wp/heart-failure-guidelines-2/.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section*** *1a.7*

☐ No → *report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist,* ***provide what is known from the guideline review of evidence in*** *1a.7*

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section *1a.7*

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

2013 ACCF/AHA Guideline for the Management of Heart Failure

This guideline covers multiple management issues for the adult patient with Heart Failure (HF) including the guideline-directed medical therapy (GDMT) such as the combination of hydralazine and isosorbide dinitrate for African American patients receiving ACE/ARB therapy.

HFSA 2010 Comprehensive Heart Failure Practice Guideline

The guideline developed by HFSA in 2010 addresses prevention, evaluation, disease management and therapies (pharmacologic and device) and end of life management.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

2013 ACCF/AHA Guideline for the Management of Heart Failure

An overall grade for the quality of evidence was not assigned. Rather, the quality of a study (or set of studies) supporting a recommendation was graded on an estimate of the certainty or precision of the treatment effect.

The recommendation to support this measure is Level of Evidence of A: Data derived from multiple randomized clinical trials or meta- analyses. References used to determine level of evidence must be provided and cited with the recommendation.

HFSA 2010 Comprehensive Heart Failure Practice Guideline

The recommendations from this guideline in support of the measure are Strength of Evidence:

- A: Randomized, Controlled, Clinical Trials; May be assigned based on results of a single methodologically rigorous trial and
- B: Cohort and Case-Control Studies Post hoc, subgroup analysis, and meta-analysis; Prospective observational studies or registries

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

2013 ACCF/AHA Guideline for the Management of Heart Failure

Level of Evidence (LOE) is an estimate of the certainty or precision of the treatment effect.

Level of Evidence of 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. References used to determine level of evidence must be provided and cited with the recommendation.

Level of Evidence C: Consensus opinion of experts, case studies, or standard of care.

HFSA 2010 Comprehensive Heart Failure Practice Guideline

Strength of evidence is determined both by the type of evidence available and the assessment of validity, applicability, and certainty of a specific type of evidence. Following the lead of previous guidelines, strength of evidence in this guideline is heavily dependent on the source or type of evidence used. The HFSA guideline process has used three grades (A, B, or C) to characterize the type of evidence available to support specific recommendations.

Strength of Evidence A: Randomized, Controlled, Clinical Trials; May be assigned based on results of a single methodologically rigorous trial and

Strength of evidence B: Cohort and Case-Control Studies Post hoc, subgroup analysis, and meta-analysis; Prospective observational studies or registries

Strength of Evidence C: Expert Opinion; Observational studies-epidemiologic findings; Safety reporting from large-scale use in practice

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: An extensive evidence review was conducted through October 2011 and includes selected other

references through April 2013 for the 2013 ACCF/AHA Guideline for the Management of Heart Failure. No information on the time period was provided for the HFSA 2010 Comprehensive Heart Failure Practice Guideline.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

The body of evidence supporting the recommendations on guideline-directed medical therapy includes:

4 randomized controlled trials (RCTs)

2 post hoc retrospective analyses

HFSA 2010 Comprehensive Heart Failure Practice Guideline

Specific information on the number of studies included in the body of evidence not provided.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

2013 ACCF/AHA Guideline for the Management of Heart Failure

The recommendation for this medication therapy is rated as Level of Evidence A, meaning that the data was derived from multiple RCTs or meta-analyses. Additional information on the overall quality of evidence across the RCTs is not provided.

HFSA 2010 Comprehensive Heart Failure Practice Guideline

The recommendations for hydralazine and isosorbide dinitrate therapy are rated as Strength of Evidence A and B, meaning that the data was derived from RCTs, cohort and Case-Control Studies Post hoc, subgroup analysis, meta-analysis or prospective observational studies or registries. Additional information on the overall quality of evidence is not provided.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Fonarow GC, Yancy CW, Hernandez AF, Peterson ED, Spertus JA, Heidenreich PA. Potential impact of optimal implementation of evidence-based heart failure therapies on mortality. Am Heart J. 2011;161:1024–1030.e3.

In 2011, Fonarow and colleagues complete a post hoc retrospective analysis to identify current gaps in care for patients with HF and reduced LVEF and to quantify the potential benefits of specific evidence based therapies. Review of RCT data for combination of hydralazine and isosorbide dinitrate showed that a patient's relative risk for death was reduced by 43% and the number needed to treat for mortality (standardized to 12 months) was 21. If this combination was prescribed to all of the patients for which it was appropriate, then 9.8% or 6,655 lives could be saved each year.

2013 ACCF/AHA Guideline for the Management of Heart Failure

p. e179:

In a large-scale trial that compared the vasodilator combination with placebo, the use of hydralazine and isosorbide dinitrate reduced mortality but not hospitalizations in patients with HF treated with digoxin and diuretics but not an ACE inhibitor or beta blocker. However, in 2 other trials that compared the vasodilator combination with an ACE inhibitor, the ACE inhibitor produced more favorable effects on survival. A post hoc retrospective analysis of these vasodilator trials demonstrated particular efficacy of isosorbide dinitrate and hydralazine in the African American cohort. In a subsequent trial, which was limited to patients self-described as African American, the addition of a fixed-dose combination of hydralazine and isosorbide dinitrate to standard therapy with an ACE inhibitor or ARB, a beta blocker, and an aldosterone antagonist offered significant benefit.

The combination of hydralazine and isosorbide dinitrate is recommended for African Americans with HFrEF who remain symptomatic despite concomitant use of ACE inhibitors, beta blockers, and aldosterone antagonists. Whether this benefit is evident in non–African Americans with HFrEF remains to be investigated. The combination of hydralazine and isosorbide dinitrate should not be used for the treatment of HFrEF in patients who have no prior use of standard neurohumoral antagonist therapy and should not be substituted for ACE inhibitor or ARB therapy in patients who are tolerating therapy without difficulty. Despite the lack of data with the vasodilator combination in patients who are intolerant of ACE inhibitors or ARBs, the combined use of hydralazine and isosorbide dinitrate may be considered as a therapeutic option in such patients.

HFSA 2010 Comprehensive Heart Failure Practice Guideline

p. e81:

The Vasodilator Heart Failure Trial (V-HeFT) was the first major randomized HF trial and was conducted in Veterans Administration hospitals throughout the US. Patients who remained symptomatic with mild to severe symptoms of HF despite treatment with diuretics and digoxin were randomized to a combination of hydralazine and isosorbide dinitrate or prazosin or placebo. The combination of hydralazine and isosorbide dinitrate was associated with a reduction in all-cause mortality compared to both placebo and prazosin that was of borderline statistical significance (P = .053). In V-HeFT II, the combination of hydralazine and isosorbide dinitrate was compared with enalapril in a population similar to V-HeFT I. All- cause mortality was 28% lower with enalapril than with the hydralazine isosorbide dinitrate combination. However, quality of life and peak exercise capacity as measured by peak oxygen consumption were better with hydralazine-isosorbide dinitrate.

The African-American Heart Failure Trial (A-HeFT) enrolled 1050 self-identified African-American patients

who had NYHA class III or IV HF with dilated ventricles and reduced LVEF. In this placebo-controlled, blinded, and randomized trial, subjects were randomly assigned to receive a fixed combination of isosorbide dinitrate plus hydralazine or placebo in addition to standard therapy for HF. The primary end point was a composite score made up of weighted values for death from any cause, a first hospitalization for HF, and change in the quality of life. The study was terminated early because of a significantly higher mortality rate in the placebo group than in the group given the fixed combination of isosorbide dinitrate plus hydralazine (10.2% vs 6.2%, P = .02). The mean primary composite score was significantly better in the group given isosorbide dinitrate plus hydralazine than in the placebo group, as were its individual components: 43% reduction in the rate of death from any cause, 33% relative reduction in the rate of first hospitalization for HF, and an improvement in the quality of life. These results taken together constitute a strong recommendation for the addition of the fixed combination of isosorbide dinitrate/hydralazine to the standard medical regimen for HF in African Americans. Data cannot exclude a benefit of the isosorbide dinitrate/hydralazine combination in non-African Americans when added to the standard medical regimen for HF.

p. e171:

The A-HeFT (African-American Heart Failure Trial) confirmed the benefit of hydralazine/isosorbide dinitrate in black HF patients. Importantly, 40% of the A-HeFT cohort were women. An analysis of outcomes by gender in A-HeFT showed that fixed-dose combined hydralazine/ isosorbide dinitrate improved HF outcomes in both men and women. There were no gender differences between men and women in the benefit of hydralazine/isosorbide dinitrate on the primary composite score, time to first HF hospitalization, and event-free survival.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

2013 ACCF/AHA Guideline for the Management of Heart Failure

p. e179:

Adherence to this combination has generally been poor because of the large number of tablets required, frequency of administration, and the high incidence of adverse reactions. Frequent adverse effects include headache, dizziness, and gastrointestinal complaints. Nevertheless, the benefit of these drugs can be substantial and warrant a slower titration of the drugs to enhance tolerance of the therapy.

HFSA 2010 Comprehensive Heart Failure Practice Guideline

Potential harms were not addressed in this review of the evidence.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Two additional analyses from A-HeFT were published after the publication of the 2013 ACCF/AHA Guideline for the Management of Heart Failure.

Note: Text below for description and results is verbatim from the article abstract.

Anand IS, Win S, Rector TS, Cohn JN, Taylor AL. Effect of fixed-dose combination of isosorbide dinitrate and hydralazine on all hospitalizations and on 30-day readmission rates in patients with heart failure: results from the African-American Heart Failure Trial. Circ Heart Fail. 2014;7:759-65. doi: 10.1161/CIRCHEARTFAILURE.114.001360. Epub 2014 Jun 26.

Background: Fixed-dose combination of isosorbide dinitrate and hydralazine (FDC-I/H) reduced mortality by 43% and death or first hospitalization for heart failure (HF) by 37% in the African-American Heart Failure Trial (A-HeFT). Reduction in mortality makes it difficult to determine the effect on hospitalizations unless the analysis adjusts for death as a competing risk.

Methods and Results: In A-HeFT, 1050 self-identified black patients with moderate to severe HF were randomized to FDC-I/H or placebo. The effects of FDC-I/H on first and all hospitalizations and 30-day readmission rates were analyzed. Deaths as competing risks were adjusted using Fine-Gray regression and joint models of hospitalizations and mortality. There were 558 all-cause and 251 HF hospitalizations in placebo compared with 435 and 173 hospitalizations in the FDC-I/H group. Adjusting for deaths as a competing risk, the effect of FDC-I/H on the first hospitalization for HF, expressed in hazard ratio (95% confidence interval), was 0.61 (0.47-0.80; P<0.001) and 0.88 (0.72-1.06; P=0.18) on the first all-cause hospitalization. The effect of FDC-I/H on all recurrent hospitalizations for HF was 0.66 (0.52-0.83; P=0.0005), similar to the effect on the first hospitalizations for HF, whereas the effect on all hospitalizations for any cause was 0.75 (0.63-0.91; P=0.003). The 30-day all-cause readmission rate after the first hospitalization for HF was 23.6% (29 of 123) in placebo versus 14.8% (12 of 81) in the FDC-I/H group, but the effect (0.59; 0.30-1.16; P=0.12) in this small subgroup was not significant.

Conclusions: Treatment with FDC-I/H was associated with a substantial reduction in the first and recurrent HF hospitalizations, and in total all-cause hospitalizations, reducing the total burden of costly and distressing hospitalizations.

McNamara DM, Taylor AL, Tam SW, Worcel M, Yancy CW, Hanley-Yanez K, Cohn JN, Feldman AM. G-protein beta-3 subunit genotype predicts enhanced benefit of fixed-dose isosorbide dinitrate and hydralazine: results of A-HeFT. JACC Heart Fail. 2014;2:551-7. doi: 10.1016/j.jchf.2014.04.016. Epub 2014 Oct 8.

Objectives: The purpose of this study was to evaluate the influence of the guanine nucleotide-binding proteins (G-proteins), beta-3 subunit (GNB3) genotype on the effectiveness of a fixed-dose combination of isosorbide dinitrate and hydralazine (FDC I/H) in A-HeFT (African American Heart Failure Trial).

Background: GNB3 plays a role in alpha2-adrenergic signaling. A polymorphism (C825T) exists, and the T allele is linked to enhanced alpha-adrenergic tone and is more prevalent in African Americans.

Methods: A total of 350 subjects enrolled in the genetic substudy (GRAHF [Genetic Risk Assessment of Heart Failure in African Americans]) were genotyped for the C825T polymorphism. The impact of FDC I/H on a composite score (CS) that incorporated death, hospital stay for heart failure, and change in quality of life (QoL) and on event-free survival were assessed in GNB3 genotype subsets.

Results: The GRAHF cohort was 60% male, 25% ischemic, 97% New York Heart Association functional class III, age 57 ± 13 years, with a mean qualifying left ventricular ejection fraction of 0.24 ± 0.06. For GNB3 genotype, 184 subjects were TT (53%), 137 (39%) CT, and 29 (8%) were CC. In GNB3 TT subjects, FDC I/H improved the CS (FDC I/H = 0.50 ± 1.6; placebo = -0.11 ± 1.8, p = 0.02), QoL (FDC I/H = 0.69 ± 1.4; placebo = 0.24 ± 1.5, p = 0.04), and event-free survival (hazard ratio: 0.51, p = 0.047), but not in subjects with the C allele (for CS, FDC I/H = -0.05 ± 1.7; placebo = -0.09 ± 1.7, p = 0.87; for QoL, FDC I/H = 0.28 ± 1.5; placebo = 0.14 ± 1.5, p = 0.56; and for event-free survival, p = 0.35).

Conclusions: The GNB3 TT genotype was associated with greater therapeutic effect of FDC I/H in A-HeFT. The role of the GNB3 genotype for targeting therapy with FDC I/H deserves further study.

Impact on conclusions of systematic review: While additional research on whether use of hydralazine and isosorbide dinitrate is linked to a genetic polymorphism may refine the clinical recommendations, findings in these publications further support the current recommendations and level of evidence ratings for the use of combination therapy in African American patients.

1a.8 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.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*
The African-American Heart Failure Trial (A-HeFT) first published in 2004 demonstrated that there is significant benefit for African American patients who receive the fixed-dose combination therapy of hydralazine and isosorbide dinitrate. A-HeFT built on the findings from the two Vasodilator-Heart Failure Trials (V-HeFT). A-HeFT, which was ended early due to the mortality rates in the placebo population, demonstrated a 43% reduction in mortality, a 33% decrease in initial hospitalizations, and a 50% improvement in patient-reported quality of life (Taylor, 2004; Sharma, 2014). These results clearly demonstrate that the fixed-dose combination therapy significantly improves patient morbidity, mortality and quality of life in this clinical cohort. There is no substitute for the fixed-dose combination therapy.

Even with this strong evidence of unprecedented efficacy and cost-effectiveness, research shows that more than 85% of African American patients are not receiving the quality of care that this therapy affords, constituting a significant gap in care quality (Dickson, 2015). The underuse of the fixed-dose combination of hydralazine plus isosorbide dinitrate in African Americans with severe heart failure is a health care and health quality disparity that exposes these patients to an elevated risk for mortality and hospitalization, and compromises efforts to contain the escalating system costs by preventing or reducing unnecessary hospitalizations and readmissions.

Based upon research on the mortality benefit of the fixed-dose combination (Fonarow, 2011), the National Minority Quality Forum estimates that 51,542 (27%) of the 189,891 African American Medicare beneficiaries who were being treated for heart failure and received their prescription drugs under Part D should have been treated with the fixed-dose combination; but only 2,377 (5%) had at least one prescription (30-day supply) of the therapy. Further, the National Minority Quality Forum estimates that between 2008 and 2010, only 3% of the eligible patient cohort in Medicare received the therapy. Given the documented number to treat to receive the

mortality benefit (21), it can be estimated that from 2007 through 2010, 20,000 African American Medicare beneficiaries died as a result of the failure to receive quality care as defined by evidence-based guidelines.

Research continues to explore if the fixed-dose combination of hydralazine and isosorbide dinitrate is linked to a particular genetic polymorphism (NIH funded Genomic Response Analysis of Heart Failure Therapy in African Americans). While we anticipate that the evidence supporting this treatment will be refined over time, the proven benefits to this patient population is significant and there is a clear opportunity for improvement. Failure to do so constitutes a failure to provide quality and cost-effective care.

References:

Dickson VV, Knafl GJ, Wald J, Riegel B. Racial differences in clinical treatment and self-care behaviors of adults with chronic heart failure. J Am Heart Assoc. 2015;4:1-13.
Fonarow GC, Yancy CW, Hernandez AF, Peterson ED, Spertus JA, Heidenreich PA. Potential impact of optimal implementation of evidence-based heart failure therapies on mortality. Am Heart J. 2011;161:1024-1030.
Sharma A, Colvin-Adams M, Yancy CW. Heart failure in African Americans: disparities can be overcome. Cleve Clin J Med. 2014;81:301-11.
Taylor AL, Ziesche S, Yancy C, et al. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. N Engl J Med 2004; 351:2049–57.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
As this is a newly developed measure, we do not yet have data on the overall performance of the measure but will be able to submit this information at the time of maintenance.

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.
Several analyses on whether eligible patients are receiving the hydralazine and isosorbide dinitrate combination therapy as supported by current evidence have been published. All demonstrate the existence of a significant opportunity for improvement both in the ambulatory setting and at the time of discharge from a hospital.

- A secondary analysis of data identified that more than 85% of African American patients were not receiving the combination therapy (Dickson, 2015).
- An observational analysis of data from the Get With the Guidelines-Heart Failure Registry showed that just over 22% of African American patients were discharged from the hospital with a prescription for the combination therapy. Rates did increase from 16% to 24% over four years (Golwala, 2013).
- A post hoc retrospective analysis conducted by Fonarow and colleagues using data from IMPROVE HF and Get with the Guidelines registry identified that only 7.3% of African American patients received the recommended combination therapy of hydralazine and isosorbide dinitrate (Fonarow, 2011).
- Rates are similarly low in the outpatient setting with the IMPROVE-HF, a prospective cohort study, showing that only 7.3% of patients received hydralazine and isosorbide dinitrate (Yancy, 2010).
- Only 4.5% of African American patients with HF and LVSD included in the OPTIMIZE-HF registry received the combination therapy (Yancy, 2008).

References:

Dickson VV, Knafl GJ, Wald J, Riegel B. Racial differences in clinical treatment and self-care behaviors of adults with chronic heart failure. J Am Heart Assoc. 2015;4:1-13.
Fonarow GC, Yancy CW, Hernandez AF, Peterson ED, Spertus JA, Heidenreich PA. Potential impact of optimal implementation of evidence-based heart failure therapies on mortality. Am Heart J. 2011;161:1024-1030.
Golwala HB, et al. Use of hydralazine-isosorbide dinitrate combination in African American and other race/ethnic group patients with heart failure and reduced left ventricular ejection fraction. J Am Heart Assoc. 2013;2:e000214. doi: 10.1161/JAHA.113.000214.
Yancy CW, Fonarow GC, Albert NM, Curtis AB, Stough WG, Gheorghiade M, Heywood JT, McBride ML, Mehra MR, O´Connor CM, Reynolds D, Walsh MN. Adherence to guideline-recommended adjunctive heart failure therapies among outpatient cardiology practices (findings from IMPROVE HF). Am J Cardiol. 2010;105:255–260.

Yancy CW, Abraham WT, Albert NM, Clare R, Stough WG, Gheorghiade M, Greenberg BH, O´Connor CM, She L, Sun JL, Young JB, Fonarow GC. Quality of care of and outcomes for African Americans hospitalized with heart failure: findings from the OPTIMIZE-HF (Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients With Heart Failure) registry. J Am Coll Cardiol. 2008;51:1675–1684.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
As this is a newly developed measure, we do not yet have data on the overall performance of the measure but will be able to submit this information at the time of maintenance.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Heart failure is a major public health burden in the United States that disproportionately affects African Americans, who have not experienced the same benefit from treatment as white patients have. More than 5 million people = 20 years of age in the United States have heart failure, with 550,000 new cases of heart failure diagnosed each year. In the US, heart failure affects about 3% of the African American populations; whereas this rate is about 2% in the general population (Ferdinand, 2014). According to the American Heart Association heart disease and stroke statistics 2014 update, annual rates per 1,000 population of new heart failure events are 16.9 and 25.5 for Black men aged 65-74 and 75-84, respectively; and 14.2 and 25.5 for Black women aged 65-74 and 75-84, respectively (Go, 2014).

Heart failure is more prevalent in African Americans than in whites, occurs earlier, imposes higher rates of death and morbidity, and has a more malignant course. Much of the disparity can be assigned to modifiable risk factors such as uncontrolled hypertension and on suboptimal health care. Therefore, when African Americans are treated according to guidelines, discrepant outcomes can be minimized (Sharma, 2014). According to American Heart Association statistics, the annual incidence of heart failure in whites is approximately 6 per 1,000 person-years, while in African Americans it is 9.1 per 1,000 person years. In the Atherosclerosis Risk in Communities Study, the incidence of new heart failure as 1.0 per 1,000 person-years in Chinese Americans, 2.4 in whites, 3.5 in Hispanics, and 4.6 in African Americans. Moreover, when hospitalized for heart failure, African Americans have a 45% greater risk of death or decline in functional status than whites. In the Women's Health Initiative — a 15-year study initiated by the National Institutes of Health in 1991 — African American women had higher rates of heart failure than white women, possibly linked to higher rates of diabetes (Sharma, 2014).

This measure specifically targets Black or African American patients with heart failure and left ventricular systolic dysfunction where a specific therapy is supported by evidence-based guidelines. For this reason, the data provided here are identical to 1b.3.

Several published analyses on whether eligible patients are receiving the hydralazine and isosorbide dinitrate combination therapy as supported by current evidence are highlighted below. All demonstrate the existence of a significant opportunity for improvement both in the ambulatory setting and at the time of discharge from a hospital.

- A secondary analysis of data identified that more than 85% of African American patients were not receiving the combination therapy (Dickson, 2015).
- An observational analysis of data from the Get With the Guidelines-Heart Failure Registry showed that just over 22% of African American patients were discharged from the hospital with a prescription for the combination therapy. Rates did increase from 16% to 24% over four years (Golwala, 2013).
- A post hoc retrospective analysis conducted by Fonarow and colleagues using data from IMPROVE HF and Get with the Guidelines registry identified that only 7.3% of African American patients received the recommended combination therapy of hydralazine and isosorbide dinitrate (Fonarow, 2011).
- Rates are similarly low in the outpatient setting with the IMPROVE-HF, a prospective cohort study, showing that only 7.3% of patients received hydralazine and isosorbide dinitrate (Yancy, 2010).
- Only 4.5% of African American patients with HF and LVSD included in the OPTIMIZE-HF registry received the combination therapy (Yancy, 2008).

References:

Dickson VV, Knafl GJ, Wald J, Riegel B. Racial differences in clinical treatment and self-care behaviors of adults with chronic heart failure. J Am Heart Assoc. 2015; 4:1-13.

Ferdinand K. Customizing therapy for African Americans with Heart Failure: Improving Outcomes and Reducing Readmissions. A CME-certified Grand Rounds Activity. Rockpoint 2014.
Fonarow GC, Yancy CW, Hernandez AF, Peterson ED, Spertus JA, Heidenreich PA. Potential impact of optimal implementation of evidence-based heart failure therapies on mortality. Am Heart J. 2011;161:1024-1030.
Golwala HB, et al. Use of hydralazine-isosorbide dinitrate combination in African American and other race/ethnic group patients with heart failure and reduced left ventricular ejection fraction. J Am Heart Assoc. 2013;2:e000214. DOI: 10.1161/JAHA.113.000214.
Go AS, et al. Heart Disease and Stroke Statistics—2014 Update: A Report From the American Heart Association Statistics committee and Stroke Statistics Subcommittee. Circulation. 2014;129:e28–e292.
Sharma A, Colvin-Adams M, Yancy CW. Heart failure in African Americans: Disparities can be overcome. Cleveland Clinic Journal of Medicine. 2014; 81:301-311.
Yancy CW, Fonarow GC, Albert NM, Curtis AB, Stough WG, Gheorghiade M, Heywood JT, McBride ML, Mehra MR, O´Connor CM, Reynolds D, Walsh MN. Adherence to guideline-recommended adjunctive heart failure therapies among outpatient cardiology practices (findings from IMPROVE HF). Am J Cardiol. 2010;105:255–260.
Yancy CW, Abraham WT, Albert NM, Clare R, Stough WG, Gheorghiade M, Greenberg BH, O´Connor CM, She L, Sun JL, Young JB, Fonarow GC. Quality of care of and outcomes for African Americans hospitalized with heart failure: findings from the OPTIMIZE-HF (Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients With Heart Failure) registry. J Am Coll Cardiol. 2008;51:1675–1684.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

A leading cause of morbidity/mortality, Patient/societal consequences of poor quality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

This measure specifically relates to the National Quality Strategy (NQS) priority area of Effective Clinical Care: Promoting the most effective prevention and treatment practices for the leading causes of mortality, starting with cardiovascular disease.

Heart failure is a major public health burden in the United States that disproportionately affects African Americans, who have not experienced the same benefit from treatment as white patients have.

- More than 5 million people = 20 years of age in the United States have heart failure, with 550,000 new cases of heart failure diagnosed each year (Ferdinand, 2014).
- In the US, heart failure affects about 3% of the African American populations; whereas, this rate is about 2% in the general population (Ferdinand, 2014).
- According to the American Heart Association (AHA) heart disease and stroke statistics 2014 update, annual rates per 1,000 population of new heart failure events are 16.9 and 25.5 for Black men aged 65-74 and 75-84, respectively; and 14.2 and 25.5 for Black women aged 65-74 and 75-84, respectively (Go, 2014).

Heart failure is more prevalent in African Americans than in whites, occurs earlier, imposes higher rates of death and morbidity, and has a more malignant course. Much of the disparity can be assigned to modifiable risk factors such as uncontrolled hypertension and on suboptimal health care. Therefore, when African Americans are treated according to guidelines, discrepant outcomes can be minimized (Sharma, 2014).

- According to AHA statistics, the annual incidence of heart failure in whites is approximately 6 per 1,000 person-years, while in African Americans it is 9.1 per 1,000 person years.
- In the Atherosclerosis Risk in Communities Study, the incidence of new heart failure as 1.0 per 1,000 person-years in Chinese Americans, 2.4 in whites, 3.5 in Hispanics, and 4.6 in African Americans. Moreover, when hospitalized for heart failure, African Americans have a 45% greater risk of death or decline in functional status than whites.
- In the Women's Health Initiative — a 15-year study initiated by the National Institutes of Health in 1991 — African American women had higher rates of heart failure than white women, possibly linked to higher rates of diabetes (Sharma, 2014).

The African-American Heart Failure Trial (A-HeFT) first published in 2004 demonstrated that there is significant benefit for African American patients who receive the fixed-dose combination therapy of hydralazine and isosorbide dinitrate. A-HeFT built on the findings from the two Vasodilator-Heart Failure Trials (V-HeFT). A-HeFT, which was ended early due to the mortality rates in the placebo population, demonstrated a 43% reduction in mortality, a 33% decrease in initial hospitalizations, and a 50% improvement in patient-reported quality of life (Taylor, 2004; Sharma, 2014). These results clearly demonstrate that the fixed-dose combination therapy significantly improves patient morbidity, mortality and quality of life in this clinical cohort. There is no substitute for the fixed-dose combination therapy.

Based upon research on the mortality benefit of the fixed-dose combination (Fonarow, 2011), the National Minority Quality Forum estimates that 51,542 (27%) of the 189,891 African American Medicare beneficiaries who were being treated for heart failure and received their prescription drugs under Part D should have been treated with the fixed-dose combination; but only 2,377 (5%) had at least one prescription (30-day supply) of the therapy. Further, the National Minority Quality Forum estimates that between 2008 and 2010, only 3% of the eligible patient cohort in Medicare received the therapy. Given the documented number to treat to receive the mortality benefit (21), it can be estimated that from 2007 through 2010, 20,000 African American Medicare beneficiaries died as a result of the failure to receive quality care as defined by evidence-based guidelines.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
References:

Ferdinand K. Customizing therapy for African Americans with Heart Failure: Improving Outcomes and Reducing Readmissions. A CME-certified Grand Rounds Activity. Rockpoint 2014.
Fonarow GC, Yancy CW, Hernandez AF, Peterson ED, Spertus JA, Heidenreich PA. Potential impact of optimal implementation of evidence-based heart failure therapies on mortality. Am Heart J. 2011;161:1024-1030.
Go AS, et al. Heart Disease and Stroke Statistics—2014 Update: A Report From the American Heart Association Statistics committee and Stroke Statistics Subcommittee. Circulation. 2014;129:e28–e292.
Sharma A, Colvin-Adams M, Yancy CW. Heart failure in African Americans: Disparities can be overcome. Cleveland Clinic Journal of Medicine. 2014; 81:301-311.
Taylor AL, Ziesche S, Yancy C, et al. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. N Engl J Med 2004; 351:2049–57.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable

Testing attachment

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria 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 : Congestive Heart Failure

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Disparities, Safety

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.)*
http://heartfailurequalityimprovementinitiative.com/performance-measures/

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:** NMQF_fixed_dose_thrpy_Bonnie_test_data.zip,NMQF_fixed_dose_thrpy_eMeasure_final.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:** NMQF_fixed_dose_thrpy_value_sets.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable

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)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients prescribed a fixed-dose combination of hydralazine and isosorbide dinitrate seen for an office visit in the measurement period in the outpatient setting or at each hospital discharge

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Measurement period (12 months)

S.6. **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, 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.
The following data element is used to calculate the numerator:
1. Fixed-dose combination of hydralazine and isosorbide dinitrate prescription

Logic for calculating the numerator is included in the eMeasure specification.

Value sets used:
Fixed dose combination of hydralazine and isosorbide dinitrate (2.16.840.1.113762.1.4.1124.15)

S.7. **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 EF <40% who are self-identified Black or African Americans and receiving ACEI or ARB and Beta-blocker therapy

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The following data elements are used to calculate the denominator:
1. Diagnosis of heart failure
2. Ejection Fraction <40% or diagnosis of left ventricular systolic dysfunction
3. Self-identified as Black or African American
4. ACEI or ARB therapy

5. Beta-blocker therapy
6. Office visit
7. Hospital Discharge

Logic for calculating the denominator is included in the eMeasure specification.

Value sets used:
Heart Failure (2.16.840.1.113883.3.526.2.23, 2.16.840.1.113883.3.526.2.24, 2.16.840.1.113883.3.526.2.25, 2.16.840.1.113883.3.526.3.376)
Left Ventricular Systolic Dysfunction (2.16.840.1.113883.3.526.2.859, 2.16.840.1.113883.3.526.3.1091)
Moderate or Severe LVSD (2.16.840.1.113883.3.526.2.861, 2.16.840.1.113883.3.526.3.1090)
Ejection Fraction (2.16.840.1.113883.3.526.2.1238, 2.16.840.1.113883.3.526.3.1134)
Moderate or Severe (2.16.840.1.113883.3.526.3.1092)
Care Services in Long-Term Residential Facility (2.16.840.1.113883.3.464.1003.101.11.1070, 2.16.840.1.113883.3.464.1003.101.12.1014)
Self identified as Black or African American (2.16.840.1.113762.1.4.1124.1)
Discharge Services - Hospital Inpatient (2.16.840.1.113883.3.464.1003.101.11.1035, 2.16.840.1.113883.3.464.1003.101.12.1007)
Face-to-Face Interaction (2.16.840.1.113883.3.464.1003.101.11.1216, 2.16.840.1.113883.3.464.1003.101.12.1048)
Home Healthcare Services (2.16.840.1.113883.3.464.1003.101.11.1080, 2.16.840.1.113883.3.464.1003.101.12.1016)
Nursing Facility Visit (2.16.840.1.113883.3.464.1003.101.11.1060, 2.16.840.1.113883.3.464.1003.101.12.1012)
Office Visit (2.16.840.1.113883.3.464.1003.101.11.1005, 2.16.840.1.113883.3.464.1003.101.12.1001)
Outpatient Consultation (2.16.840.1.113883.3.464.1003.101.11.1040, 2.16.840.1.113883.3.464.1003.101.12.1008)
Patient provider interaction (2.16.840.1.113883.3.526.2.1049, 2.16.840.1.113883.3.526.3.1012)
ACE Inhibitor or ARB (2.16.840.1.113883.3.526.2.39, 2.16.840.1.113883.3.526.3.1139)
Beta Blocker Therapy for LVSD (2.16.840.1.113883.3.526.2.133, 2.16.840.1.113883.3.526.3.1174)

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Denominator exclusions include:
- Hypotension (severe or symptomatic)
- Severe lupus erythematosus
- Unstable angina
- Peripheral neuritis
- Patient actively taking Phosphodiesterase Type 5 (PDE5) Inhibitors

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The following data elements are used to calculate the denominator exclusions:
1. Hypotension (severe or symptomatic)
2. Severe lupus erythematosus
3. Unstable angina
4. Peripheral neuritis
5. Patient actively taking Phosphodiesterase Type 5 (PDE5) Inhibitors

Logic for calculating the denominator exclusions are included in the eMeasure specification.

Value sets used:
Hypotension (2.16.840.1.113883.3.526.2.175, 2.16.840.1.113883.3.526.2.180, 2.16.840.1.113883.3.526.2.185, 2.16.840.1.113883.3.526.3.370)
Lupus erythematosus (2.16.840.1.113762.1.4.1124.9, 2.16.840.1.113762.1.4.1124.10, 2.16.840.1.113762.1.4.1124.11, 2.16.840.1.113762.1.4.1124.12)
Unstable angina (2.16.840.1.113762.1.4.1124.16, 2.16.840.1.113762.1.4.1124.17, 2.16.840.1.113762.1.4.1124.18)
Peripheral neuritis (2.16.840.1.113762.1.4.1124.4, 2.16.840.1.113762.1.4.1124.5, 2.16.840.1.113762.1.4.1124.6, 2.16.840.1.113762.1.4.1124.7)
Patient actively taking Phosphodiesterase Type 5 (PDE5) Inhibitors (2.16.840.1.113762.1.4.1124.14)
Severe (2.16.840.1.113762.1.4.1124.19)
Symptomatic (2.16.840.1.113762.1.4.1124.20)

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*
Not applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **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.18. **Calculation Algorithm/Measure Logic** *(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; aggregating data; risk adjustment; etc.)*
The measure logic is provided in the eMeasure specification.

Performance is calculated as:

1. Identify the initial patient population for the measure.
2. From those patients in the initial patient population, identify those that meet the denominator criteria.
3. From the patients who qualify for the denominator, identify those who meet the numerator criteria.
4. Identify those patients who did not meet the numerator criteria and determine whether an appropriate exclusion is documented.
5. Remove those patients with an exclusion from the denominator.
6. Calculation: Numerator/Denominator-Denominator Exclusions

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
This measure is specified with specific criteria, data elements and value sets. If a patient record does not include one or more of these components for the initial patient population or denominator, then patients are not considered eligible for the measure and not included.

If data to determine whether a patient should be considered for the numerator or exclusions is missing, then the numerator or exclusions not considered to be met and the provider will not get credit for meeting performance for that patient.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Electronic Health Record

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Not applicable

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Hospital/Acute Care Facility
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form

National Quality Forum—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2764

Measure Title: Heart Failure: Fixed-dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black and African American Patients with Heart Failure and LVEF <40% on ACEI or ARB and Beta-blocker Therapy

Date of Submission: 8/12/2015

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process

☐ 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, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** 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 (2b2-2b6) 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 20 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 sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering 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 **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. 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 **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [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]

2b4. 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 sociodemographic 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.

2b5. 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.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), 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.

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.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/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*).

Not applicable

1.3. What are the dates of the data used in testing? Exact dates TBD but will include 12 months of performance data at a minimum to be consistent with the measure specifications.

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:	**Measure Tested at Level of:**

(*must be consistent with levels entered in item S.26*)	
☒ 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*)

NMQF testing will examine whether all the data elements required to calculate the performance score are correctly identified; specifically, the accuracy of the electronic data in the automated report produced from implementation of the eMeasure specifications against the manual review and abstraction of the medical record (the gold standard). NMQF will analyze and test all data elements required to capture the denominator, numerator and exclusions. We will not only look to determine the ability of the measure to identify positive results (agreement with the discrete fields used to obtain the electronic data within the automated report). We will also examine whether the measure can identify negative results – the data elements were missing in the automated report, yet identified in text fields (e.g., progress notes) and whether the absence of the data elements potentially negatively impacts performance scores (i.e., not captured in discrete fields).

NMQF testing will address aspects across the NQF measure evaluation criteria including:

- Reliability analyses – parallel forms (agreement between automated reports and the manual review of the medical record) and Cohen's Kappa statistic with 95 percent confidence intervals
- Validity analyses – extraction accuracy (criterion validity); sensitivity (ability of a measure to identify positive results), specificity (ability of a measure to identify negative results) and percent agreement between the extracted and gold standard data
- Usability – measures element average and a weighted average of measure elements
- Feasibility – analysis of the feasibility of EHRs to collect the needed data elements (current and future) through site interviews with both clinical and informatics staff at each testing site

To power the analysis for statistical significance, the sample sizes required are estimated to be between 165 and 200 patients (given 80 percent power, and a 0.05 significance level for testing differences in Kappa statistics and performance rates). To yield an adequate sample size, an abstractor will review an estimated average of 55 medical records for each of the anticipated three to four sites in the cohort. Larger sites could provide larger samples of cases and smaller sites provide fewer; yet, all will provide a representative sample of eligible cases seen at their sites.

NMQF intends to include more than one electronic health record system (EHRs) in the testing of this measure as specified by the NQF Measure Evaluation Criteria for eMeasures. Because the measure examines the care provided at the time of discharge from a hospital, and routine care in an ambulatory practice, both settings will

be included in the testing and analyses. NMQF will also seek to include testing sites of various sizes and geographic locations to the greatest extent possible to ensure representative results.

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

A random sample of patients meeting the measure specifications will be used in the testing and analysis. To power the analysis for statistical significance, the sample sizes required are estimated to be between 165 and 200 patients (given 80 percent power, and a 0.05 significance level for testing differences in Kappa statistics and performance rates). To yield an adequate sample size, an abstractor will review an estimated average of 55 medical records for each of the anticipated three to four practice sites in the cohort. Larger sites could provide larger samples of cases and smaller sites provide fewer; yet, all will provide a representative sample of eligible cases seen at their practice sites.

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.

One data set will be produced from the hospitals and ambulatory practices identified in 1.5; therefore, no differences in the data or samples across the various aspects of testing are anticipated.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Not applicable

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*)

NMQF testing will examine whether all the data elements required to calculate the performance score are correctly identified; specifically, the accuracy of the electronic data in the automated report produced from implementation of the eMeasure specifications against the manual review and abstraction of the medical record (the gold standard). NMQF will analyze and test all data elements required to capture the denominator, numerator and exclusions. We will not only look to determine the ability of the measure to identify positive results (agreement with the discrete fields used to obtain the electronic data within the automated report). We will also examine whether the measure can identify negative results – the data elements were missing in the automated report, yet identified in text fields (e.g., progress notes) and whether the absence of the data elements potentially negatively impacts performance scores (i.e., not captured in discrete fields).

The data analysis will provide results (overall and by denominator, numerator and exclusions) on the:

- Reliability analyses – parallel forms (agreement between automated reports and the manual review of the medical record) and Cohen's Kappa statistic with 95 percent confidence intervals
- Validity analyses – extraction accuracy (criterion validity); sensitivity (ability of a measure to identify positive results), specificity (ability of a measure to identify negative results) and percent agreement between the extracted and gold standard data
- Usability – measure element average and a weighted average of measure elements
- Feasibility – analysis of the feasibility of EHRs to collect the needed data elements through site interviews with both clinical and informatics staff at each testing site

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*)

To be provided when testing is completed

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?*)

To be provided when testing is completed

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

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

☐ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. 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)*

NMQF testing will examine whether all the data elements required to calculate the performance score are correctly identified; specifically, the accuracy of the electronic data in the automated report produced from implementation of the eMeasure specifications against the manual review and abstraction of the medical record (the gold standard). NMQF will analyze and test all data elements required to capture the denominator, numerator and exclusions. We will not only look to determine the ability of the measure to identify positive results (agreement with the discrete fields used to obtain the electronic data within the automated report). We will also examine whether the measure can identify negative results – the data elements were missing in the automated report, yet identified in text fields (e.g., progress notes) and whether the absence of the data elements potentially negatively impacts performance scores (i.e., not captured in discrete fields).

The data analysis will provide results (overall and by denominator, numerator and exclusions) on the:

- Reliability analyses – parallel forms (agreement between automated reports and the manual review of the medical record) and Cohen's Kappa statistic with 95 percent confidence intervals
- Validity analyses – extraction accuracy (criterion validity); sensitivity (ability of a measure to identify positive results), specificity (ability of a measure to identify negative results) and percent agreement between the extracted and gold standard data
- Usability – measure element average and a weighted average of measure elements
- Feasibility – analysis of the feasibility of EHRs to collect the needed data elements through site interviews with both clinical and informatics staff at each testing site

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

To be provided when testing is completed

2b2.4. 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?*)

To be provided when testing is completed

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

NMQF will examine the overall frequency of the exclusions as well as variability across the testing sites to demonstrate the need for exclusions. As discussed in 2b2.2, NMQF testing will examine whether all the data elements required to calculate the performance score are correctly identified; specifically, the accuracy of the electronic data in the automated report produced from implementation of the eMeasure specifications against the manual review and abstraction of the medical record (the gold standard). NMQF will analyze and test all data elements required to capture the denominator, numerator and exclusions. We will not only look to determine the ability of the measure to identify positive results (agreement with the discrete fields used to

obtain the electronic data within the automated report). We will also examine whether the measure can identify negative results – the data elements were missing in the automated report, yet identified in text fields (e.g., progress notes) and whether the absence of the data elements potentially negatively impacts performance scores (i.e., not captured in discrete fields).

The data analysis will provide results (overall and by denominator, numerator and exclusions) on the:

- Reliability analyses – parallel forms (agreement between automated reports and the manual review of the medical record) and Cohen's Kappa statistic with 95 percent confidence intervals
- Validity analyses – extraction accuracy (criterion validity); sensitivity (ability of a measure to identify positive results), specificity (ability of a measure to identify negative results) and percent agreement between the extracted and gold standard data
- Usability – measure element average and a weighted average of measure elements
- Feasibility – analysis of the feasibility of EHRs to collect the needed data elements through site interviews with both clinical and informatics staff at each testing site

2b3.2. 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*)

To be provided when testing is completed

2b3.3. 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*)

To be provided when testing is completed

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

Not applicable

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

Not applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

Not applicable

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. 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*)

Not applicable

2b4.11. 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*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. 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)*

NMQF will examine the variation in performance scores across the testing sites. Specifically, we will look at the mean, standard deviations, and range (minimum and maximum) levels of performance to determine if there is room for improvement and meaningful differences in performance.

NMQF will initially classify a provider or site as high or low performing using a known or proven measurement of quality. Then a discriminant analysis will be performed to determine the total probability of misclassification using the new measure. If the underlying assumption of a multivariate distribution is not met, a logistic regression may be performed instead of a discriminant analysis. This logistic regression will also require an initial classification as low or high performing using a known/proven measurement of quality. The result of the logistic regression will determine the probability that a provider belongs to a particular category.

Furthermore, a chi square test will examine whether the proportion of the high performing group with a specific score is different from the proportion of the low performing group with a specific score (i.e. is there a difference between high and low performing scoring 80% or greater on the new measure).

2b5.2. 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*)

To be provided when testing is completed

2b5.3. 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?*)

To be provided when testing is completed

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i*.e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

As discussed in 2b2.2, NMQF testing will examine whether all the data elements required to calculate the performance score are correctly identified; specifically, the accuracy of the electronic data in the automated report produced from implementation of the eMeasure specifications against the manual review and abstraction of the medical record (the gold standard). NMQF will analyze and test all data elements required to capture the denominator, numerator and exclusions. We will not only look to determine the ability of the measure to identify positive results (agreement with the discrete fields used to obtain the electronic data within the automated report). We will also examine whether the measure can identify negative results – the data elements were missing in the automated report, yet identified in text fields (e.g., progress notes) and whether the absence of the data elements potentially negatively impacts performance scores (i.e., not captured in discrete fields).

The data analysis will provide results (overall and by denominator, numerator and exclusions) on the:

- Reliability analyses – parallel forms (agreement between automated reports and the manual review of the medical record) and Cohen's Kappa statistic with 95 percent confidence intervals
- Validity analyses – extraction accuracy (criterion validity); sensitivity (ability of a measure to identify positive results), specificity (ability of a measure to identify negative results) and percent agreement between the extracted and gold standard data
- Usability – measure element average and a weighted average of measure elements
- Feasibility – analysis of the feasibility of EHRs to collect the needed data elements through site interviews with both clinical and informatics staff at each testing site

2b7.2. 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*)

To be provided when testing is completed

2b7.3. 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*)

To be provided when testing is completed

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)

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)*
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.**

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.**
Attachment **Attachment:** National_Minority_Quality_Forum_Feasibility_Assessment_of_Fixed_Dose_Therapy_Measure.pdf

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Because this measure is submitted for eMeasure trial approval and testing is not yet completed, we are not yet able to share information on data availability and collection beyond what is provided in the feasibility assessment in 3b.3. This information will be collected during our testing and modifications made to the measure based on the results.

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)*.
Not applicable

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.

Planned	Current Use (for current use provide URL)
Payment Program	Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

	Get with the Guideline-Heart Failure Registry http://www.heart.org/HEARTORG/HealthcareResearch/GetWithTheGuidelines/GetWithTheGuidelines-HF/Get-With-The-Guidelines-Heart-Failure_UCM_306087_SubHomePage.jsp

4a.1. **For each CURRENT use, checked above, provide:**

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

A similar measure focused on hospital performance is currently used for quality improvement and benchmarking purposes in the American Heart Association´s Get with the Guidelines-Heart Failure registry. Information on the geographic area, number and percentage of hospitals, providers and patients is not available on the registry web site.

4a.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?*)

This is a newly developed measure intended to be used and reported at the clinician level. Information on additional uses including accountability applications will be provided at the time of maintenance. NMQF is dedicated to ensuring that this measure is implemented widely and submitted the measure for consideration by the Centers for Medicare & Medicaid Services (CMS) under its recent Physician Quality Reporting System (PQRS) call for measures. This measure could be considered for inclusion in CMS programs as early as 2017.

4a.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.*)

This is a newly developed measure intended to be used and reported at the clinician level. Information on additional uses including accountability applications will be provided at the time of maintenance. NMQF is dedicated to ensuring that this measure is implemented widely and submitted the measure for consideration by the Centers for Medicare & Medicaid Services (CMS) under its recent Physician Quality Reporting System (PQRS) call for measures. This measure could be considered for inclusion in CMS programs as early as 2017.

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

As this is a newly developed measure, we do not yet have data on the overall performance and progress on improvement of the measure but will be able to submit this information at the time of maintenance.

4b.2. **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.**

Because this is a newly developed measure, we do not have data on improvement other than what has been published. Specifically, Golwala and colleagues completed an observational analysis of data from the Get With the Guidelines-Heart Failure Registry. They showed that just over 22% of African American patients were discharged from the hospital with a prescription for the combination therapy. While performance would be considered low overall, rates increased from 16% to 24% over four years (Golwala, 2013). NMQF is also aware of individual providers and hospitals who are tracking this data through the Get with the Guidelines-Heart Failure registry and actively working toward improving performance on this measure. This measure is based on Class IA clinical guideline recommendations with clear reductions in mortality and initial hospitalizations and improved quality of life in patients.

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

Because this measure is submitted for eMeasure trial approval and testing is not yet completed, we are not yet able to share information on whether unintended negative consequences were identified. This information will be collected during our testing and modifications and other actions to mitigate them will be made to the measure based on the results.

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)**

0081 : Heart Failure (HF): Angiotensin-Converting Enzyme (ACE) Inhibitor or Angiotensin Receptor Blocker (ARB) Therapy for Left Ventricular Systolic Dysfunction (LVSD)

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

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

Previously endorsed measure:

0162: ACEI or ARB for left ventricular systolic dysfunction - Heart Failure (HF) Patients (CMS)

0610: Heart Failure - Use of ACE Inhibitor (ACEI) or Angiotensin Receptor Blocker (ARB) Therapy (ActiveHealth Management)

0615: Heart Failure - Use of Beta Blocker Therapy (ActiveHealth Management)

5a. Harmonization

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 completely harmonized?

Yes

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

Measure specifications for the target population and medication therapies for ACEI, ARB, and beta-blocker are completely harmonized with 0081 and 0083.

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.)
Not applicable

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** National Minority Quality Froum
Co.2 **Point of Contact:** Gretchen, Wartman, gwartman@nmqf.org, 202-223-7560-
Co.3 **Measure Developer if different from Measure Steward:** National Minority Quality Froum
Co.4 **Point of Contact:** Gretchen, Wartman, gwartman@nmqf.org, 202-223-7560-

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.
Writing Committee members:

Ola Akinboboye, MD, MPH, MBA, FACC, FACP
Associate Professor of Clinical Medicine, Weill Medical College of Cornell University
Medical Director, Queens Heart Institute
Rosedale, NY

Bryan Curry, MD, FACC
Department of Cardiology
Howard University Hospital
Washington, DC

Anekwe E. Onwuanyi, MD, FACC
Professor of Medicine and Chief of Cardiology, Morehouse School of Medicine
Medical Director, Heart Failure Program, Grady Health System
Atlanta, GA

David N. Smith, MD, FACP, FACC
Assistant Clinical Professor of Medicine
Yale School of Medicine
New Haven, CT

Modele O. Ogunniyi, MD, MPH, FACC, FACP, FAHA
Assistant Professor of Medicine
Division of Cardiology, Department of Medicine
Emory University School of Medicine
Assistant Medical Director, Grady Heart Failure Program
Atlanta, GA

Elizabeth Ofili, MD, MPH, FACC
Professor of Medicine
Director and Senior Associate Dean
Clinical Research Center & Clinical and Translational Research
Morehouse School of Medicine
Atlanta, GA

This committee advised on the underlying evidence, measure statements construction, and detailed specifications during the development of the Fixed-dose combination therapy measure. They will continue to provide input and clinical expertise as the measure is tested and finalized and during every measure update.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 06, 2015

Ad.4 **What is your frequency for review/update of this measure?** Specifications are updated annually; supporting guidelines reviewed 3 years or as evidence changes

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.7 **Disclaimers:** These Measures are intended to assist physicians in enhancing quality of care. Measures are designed for use by any physician who manages the care of a patient for a specific condition, or for prevention. These performance Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. NMQF encourages the testing and evaluation of its Measures.

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

Ad.8 **Additional Information/Comments:** Not applicable