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: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0133

Measure Title: In-Hospital Risk Adjusted Rate of Mortality for Patients Undergoing PCI

Measure Steward: American College of Cardiology

Brief Description of Measure: Risk adjusted rate of mortality for all patients age 18 and over undergoing PCI.

Developer Rationale: This measure allows benchmarking against the national aggregate and against hospitals with similar volume, so that hospitals with high mortality rates can engage in quality improvement to reduce mortality following PCI procedures. In-depth analysis of the causes behind variations in mortality during or post PCI can lead to the identification of best practices. Particularly actionable opportunities to improve care are to reduce peri-procedural bleeding rates and acute kidney injury, where operators have the option to pursue strategies that decrease these complications. In addition, detailed case reviews can identify operators with poorer performance for whom additional training or reduced caseloads could be considered. Active dissemination of those best practices and support to enable their adoption will improve outcomes and reduce variations in clinical practice. Improvements in the quality of care resulting from the evaluation of the risk for mortality, before and after implementing quality improvement interventions, can enable centers to quantify their improved outcomes with respect to peri-procedural mortality and a reduction in cost associated with these events. Additionally, by putting the responsibility for improved quality in the hands of physicians and other health-care practitioners, this risk-adjusted mortality measure engages the medical community around the common goal of better health-care value.

Numerator Statement: Patients 18 years of age and older with a PCI procedure performed during episode of care who expired

Denominator Statement: Patients 18 years of age and older with a PCI procedure performed during episode of care.

Denominator Exclusions: 1. NCDR Registry patients who did not have a PCI (Patient admissions with a diagnostic cath only during that admission); 2. Patient admissions with PCI who transferred to another facility on discharge

Measure Type: Outcome

Data Source: Registry Data

Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: May 09, 2007 **Most Recent Endorsement Date:** Sep 08, 2014

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary or **Summary of prior review in 2014**

The developer provided information that a "comprehensive, personalized risk assessment can lead to decreased mortality in the PCI patient population". The developer discussed how the use of guidelines, appropriate use criteria and risk models can lead to a decrease in mortality associated with PCI.

The 2014 committee acknowledged the importance of this outcome measure, noting that the importance of understanding mortality rates as a result of performance of a PCI procedure is self-evident.

Changes to evidence from last review

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

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

Updates: The developer stated that there are no updates to the evidence.

The developer provided performance data from more than 1,500 hospitals and almost 700,00 patients in 1,600 hospitals in 2015 and more than 1,600 hospitals and more than 700,00 patients demonstrating a variation in performance from 0.96% to 2.17%, respectively.

Empirical data demonstrating a relationship between the outcome to at least one healthcare process is now required. NQF guidance states that a wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.

Question for the Committee:

- *Does the stated rationale link lower mortality rates after PCI to at least one healthcare action?*
- *Is the performance data sufficient, in size and variance, to demonstrate that some hospitals are engaging in quality improvement activities to decrease mortality after PCI better than others?*

Guidance from the Evidence Algorithm

Measure assesses a health outcome (Box 1) →The relationship between the outcome and the intervention demonstrated by performance data (Box 2) →Pass

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

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

Maintenance measures – increased emphasis on gap and variation

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

The developer provided performance data from more than 1,500 hospitals and almost 700,00 patients in 1,600 hospitals in 2015 and more than 1,600 hospitals and more than 700,00 patients.

2015 Data of Mortality Adjusted Rates	2016 Data of Mortality Adjusted Rates
1: 0.96%	1: 0.92%
2: 1.23%	2: 1.2%
3: 1.42%	3: 1.41%
4:1.59%	4:1.58%
5(median): 1.76%	5(median):1.76%
6: 1.95%	6: 1.95%
7: 2.17 %	7: 2.15%
8: 2.47%	8: 2.45%
9: 3.0%	9: 2.96%

Disparities

The developer provided disparities data here. Using 2016 data from the NCDR CathPCI Registry, they performed an analysis between observed and expected mortality rates by hospital location, gender, race and insurance status. Results indicated statistically significant results, however the developers noted that after performing patient-level adjustment for mortality risk, the absolute rates were small.

Questions for the Committee:

- *Does the measure demonstrate a quality problem related to mortality in patients undergoing PCI?*
- *Is a national performance measure still warranted?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

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

RATIONALE: N/A

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

1a. Evidence:

- This outcome measure was introduced in 2007 and has been reviewed and endorsed twice since then. No change in evidence since last review in 2014.
- They provide a logic model for the outcome measure. This is a maintenance measure. No additional evidence to my knowledge.
- The measure applies directly
- This is a measure which has been available for ten years, and there are few significant differences in outcomes among hospitals The Risk adjuster is the area of greatest threat to the validity of the measure but there has been little concern raised over the past decade (of which I have heard)

- This is a maintenance measure and there is not significant update to the evidence since this was considered in 2014. The data is robust and direct with over 1500 hospitals and over 700,000 patients.
- the measure is overall minimally changed from the previously endorsed measure although major changes in the insurance ecosystem have occurred
- insurance status is not captured
- Maintenance measure - same rationale for evidence summary as 2014. This is an important outcome to monitor. Changes in results can spur a closer examination of contributing factors and quality improvement.
- pass

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

- There is still a performance gap supporting continuation of use.
- Performance data from 2015 and 2016 show rates from 0.9% to 3%. some disparities were present for hospital type, rural/urban location, sex, insurance type, and race.
- Performance data were provided and there is a gap in care--a difference in mortality that is about 3-fold, although not large in absolute terms
- Smaller opportunity for improvement: Study in the Journal of the American Heart Association (Hospital Performance on Percutaneous Coronary Intervention Process and Outcomes Measures Philip W. Chui, Craig S. Parzynski, Brahmajee K. Nallamothu, Frederick A. Masoudi, Harlan M. Krumholz, Jeptha P. Curtis, 2017) demonstrated little room for improvement except in a handful of hospitals.
- The performance gap data was presented and although small represents a doubling of risk within the variability. It does demonstrate opportunity for improvement. They did identify various subgroups to look for disparities.
- hospitals vary significantly in this outcome measure, however, specific socioeconomic data for the population (patients) and hospitals (e.g., ZIP code average income, % insured population, % underinsured, % undocumented immigrant, etc.) are not used as adjustment while these baseline differences may not be modifiable, those may be expected to influence outcomes.
- Mortality rates are pretty low and essentially unchanged for the aggregate in 2015 & 2016. Statement from measure application: "Using 2016 data from the NCDR CathPCI Registry, they performed an analysis between observed and expected mortality rates by hospital location, gender, race and insurance status. Results indicated statistically significant results, however the developers noted that after performing patient-level adjustment for mortality risk, the absolute rates were small."
- high opportunity

Criteria 2: Scientific Acceptability of Measure Properties

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: NQF Staff

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

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Evaluation of Scientific Acceptability

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.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 0133

Measure Title: In-Hospital Risk Adjusted Rate of Mortality for Patients Undergoing PCI

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? *NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*
   TIPS: Consider the following: 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?
   ☒Yes (go to Question #2)
   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.
2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?
   TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)
   ☒Yes (go to Question #4)
   ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

3. Was **empirical VALIDITY testing** of patient-level data conducted?
   ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)
   ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)
4. Was reliability testing conducted with computed performance measure scores for each measured entity?
   TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data
   ☒Yes (go to Question #5)
   ☐No (go to Question #8)
5. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? *NOTE: If multiple methods used, at least one must be appropriate.*
   TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.
   ☒Yes (go to Question #6)
   The developer performed a signal to noise analysis of 2012 data from the National Cardiovascular Data Registry (NCDR) for CathPCI Registry.
   ☐No (please explain below then go to Question #8)
6. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?
   TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?
   Do the results demonstrate sufficient reliability so that differences in performance can be identified?
   ☐High (go to Question #8)
   ☒Moderate (go to Question #8)
   The developer provided the results of the signal to noise analysis but did not provide the specific test used to analyze the data. The developer indicated that the numbers from the signal to noise analysis demonstrated variability attributable to real differences versus measurement error.
   ☐Low (please explain below then go to Question #7)

7. Was other reliability testing reported?
   ☒Yes (go to Question #8)
   ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)
8. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?
   TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)
   ☒Yes (go to Question #9)
   ☐No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?
   TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements
   Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)
   ☐Yes (go to Question #10)
   ☒No (if no, please explain below and rate Question #10 as INSUFFICIENT)
   The developer performed a test-retest reliability of the 2012 National Cardiovascular Data Registry (NCDR) for CathPCI Registry data and only provided percent agreement for key data elements. The key data elements that were analyzed included: gender; age as assessed by date of birth; cerebrovascular disease; peripheral vascular disease; chronic lung disease; prior PCI; and diabetes for the mortality risk model. The developer indicated there was no clear misclassification for the assessed data elements (<3.5% of the data elements were misclassified). However, the developer could not perform a test-retest reliability of other elements as these elements were "expected to change over time". The other data elements were: Prior cardiac arrest, GFR, NYHA classification, shock within 24 hours of PCI, indication for PCI (e.g. STEMI vs. NSTEMI vs. others), urgency of the procedure, number, appearance and location of diseased vessels, lesion severity as assessed by the SCAI definitions, BMI, and TIMI flow.
10. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?
    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?
    ☐Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)
    ☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)
    ☒Insufficient (go to Question #11)

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

1. Were all potential threats to validity that are relevant to the measure empirically assessed?
   TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

   ☒Yes (go to Question #2)

   ☐No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]
2. Analysis of potential threats to validity: Any concerns with measure exclusions?
   TIPS: Consider the following: 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)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

   ☐Yes (please explain below then go to Question #3)

   ☒No (go to Question #3)

   The developer excluded patients who transferred to another acute care facility and those that did not have a PCI performed.

   ☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)
3. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

   ☐Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

   a. Is a conceptual rationale for social risk factors included? ☒Yes ☐No

   The developer noted that given the clinical data available, social risk factors (which are not readily available) would not likely contribute much improvement to this particular risk model.

   b. Are social risk factors included in risk model? ☐Yes ☒No

   c. Any concerns regarding the risk-adjustment approach?

   *TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? 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, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

   ☐Yes (please explain below then go to Question #4)

   ☒No (go to Question #4)

   The developer presented a detailed description of the risk-adjustment method based on 40 patient clinical characteristics and not social risk factors.
4. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

☐Yes (please explain below then go to Question #5)

☒No (go to Question #5)

5. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

   ☐Yes (please explain below then go to Question #6)

   ☐No (go to Question #6)

   ☒Not applicable (go to Question #6)

6. Analysis of potential threats to validity: Any concerns regarding missing data?

   ☐Yes (please explain below then go to Question #7)

   ☒No (go to Question #7)

   The measure developer did not complete the section on Missing Data. However, the developer did describe the data quality process which noted the handling of missing data.

ASSESSMENT OF MEASURE TESTING

7. Was empirical validity testing conducted using the measure as specified and appropriate statistical test?
   Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

   ☒Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

   Empirical validity testing of the measure score was assessed by comparing the performance of the risk-adjusted model in the development sample and two validation samples.

   ☐No (please explain below then go to Question #8)

8. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?
   TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

   ☐Yes (go to Question #9)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

9. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

   ☐Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

10. Was validity testing conducted with computed performance measure scores for each measured entity?
    TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

    ☒Yes (go to Question #11)

The developer described the process of validating the risk model by performing bivariate analysis to identify candidate variables, then multivariable, logistic regression to identify clinically meaningful variables with a statistically significant association with mortality.

☐No (please explain below and go to Question #13)

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

TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

☒Yes (go to Question #12)

The developer further assessed discrimination in the model with the C-statistic. This method is used to compare the goodness of fit of logistic regression models. The developer noted the c-statistic is 0.93, which means that the probability that predicting the outcome is substantially better than chance. Models are typically considered reasonable when the C-statistic is higher than 0.7. The developer also assessed the calibration using the Hosmer-Lemeshow test (intercept = -0.00063; p=0.97) as well as the slope of the predicted versus observed risk (slope= 0.9906; p=0.097).

☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

12. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

☒High (go to Question #14)

☐Moderate (go to Question #14)

☐Low (please explain below then go to Question #13)

☐Insufficient

13. Was other validity testing reported?

☐Yes (go to Question #14)

☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

14. Was validity testing conducted with patient-level data elements?

TIPS: Prior validity studies of the same data elements may be submitted

☐Yes (go to Question #15)

☒No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

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

TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

☐Yes (go to Question #16)

☐No (please explain below and rate Question #16 as INSUFFICIENT)

16. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

☐Insufficient (go to Question #17)

17. OVERALL VALIDITY RATING

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

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

The developer provided detailed information on the analysis performed to create and validate the risk model.

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

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

2a1. Reliability-Specifications:

- I agree that reliability is moderate
- The National Cardiovascular Data Registry (NCDR) CathPCI Registry is a validated registry. It has been implemented in numerous hospitals and is reproducible
- Given that that PCI or not is not ambiguous and that death is also not a vague result it would seem that the data elements are clearly defined.
- the numerator/denominator calculation may exclude certain patients that may significantly impact outcomes:
    - patients undergoing diagnostic cath, not undergoing PCI (for whatever reason) and dying during hospitalization
    - patients undergoing diagnostic cath, then undergoing PCI, then undergoing CABG and then dying (unsure as to which dataset these deaths are counted in... the competing CABG measure)
    - patients receiving thrombolytics due to un-availitiliby of PCI (e.g., at night, weekends) and then dying during hospitalization

- patients undergoing diagnostic cath, receiving mechanical support only and dying while waiting for CABG
- patients undergoing PCI and dying whose reports are not intentionally submitted to NCDR

- Agree with staff assessment
- medium reliability

2a2. Reliability testing:

- No. Rate reliability moderate.
- signal to noise ration was only 0.7 and deteriorated with fewer procedures performed.
- No concerns
- No concerns
- No- as long as PCI properly recorded and documented there are no reliability concerns.
- see 2a1
- Agree with staff assessment
- medium

2b2. Validity testing & 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data):

- No problem with validity Rate it high.
- Validity is high without threats
- As in past reviews there is little concern over the validity testing
- No concerns
- "see 2a1 and...it is unknown if all cases are indeed submitted to NCDR or some outliers may be withheld from reporting the developer does not collect data on all cases performed in a hospital and compare to reports submitted potential for ""gaming the system"" is present in particular with the prospect of public reporting and pay-for-performance metrics"
- No concerns
- high validity

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

- high c-statistic to the model
- there are minimal exclusions and they are consistent with the evidence and with other measures
- The risk adjustment is through the American College of Cardiology tools: National Cardiovascular Data Registry (NCDR) CathPCI Registry. The above cited study did not find issues with the Risk Adjustment
- Exclusions seem appropriate
- social risk (income, insurance, ZIP code specifics, language barrier, undocumented immigrant status, etc.) are not included likely contributing to observed differences as indicated above, the competing CABG measure may include patients with a very specific risk profile excluded from this measure
- appropriate

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent
3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer noted that the data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden. Also, most data elements exist in a structured format.
- The developer provided the 2017 annual pricing for hospitals, NCDR Analytic and Reporting Services, and licensing of measure specifications which ranges from $2,900 to $50,000.
- The developer noted that 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.
- In 2014, the standing committee agreed the measure is feasible to implement, as the measure has already been in use and collected via registry with a good track record. The committee expressed concerns related to the cost of the registry and limited EMR extraction capabilities for the data elements of the measure.

Questions for the Committee:

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

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

RATIONALE: N/A

Committee Pre-evaluation Comments: Criteria 3: Feasibility

3. Feasibility:

- Feasibility has been demonstrated for ten years.
- data collection is costly for hospitals.
- The stewards report that the cath labs find the measure to be feasible.
- This has been in use of ten years so its feasibility has been demonstrated
- No concerns about feasibility except for the registry cost/EMR extraction cost. Is it possible that some facilities will be excluded from getting credit for this measure based on not having the interface/registry budget?
- "active omission of data submission of patients with unfavorable outcome may strongly influence the measure
- there is no system in place to monitor for these occurrences"
- Maintenance measure - has been collected for some time so is feasible
- high feasibility

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**
Accountability program details [Accountability program(s) – details]
This measure is currently used in two public reporting programs (Blue Distinction Centers for Cardiac Care and Quality Hospital Insight program for Anthem). The developer noted that they are delaying public reporting of this measure in the CathPCI registry (a part of the National Cardiovascular Data Registry) as they are updating to a new registry version, and until this measure can be structurally harmonized with measure #0536 (30-Day All-Cause Risk-Standardized Mortality Rate Following PCI for Patients with ST Segment Elevation Myocardial Infarction (STEMI) or Cardiogenic Shock).
4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure
Feedback on the measure by those being measured or others Feedback is obtained in a variety of ways. The developer notes that registry participants indicate that the measure is easy to understand and interpret, and that participants "seem to be apply[ing] the coding instructions correctly to the data elements that impact this measure".

Additional Feedback:

The developer mentioned previous concerns from the interventional community that have been addressed in version 4 of the CathPCI Registry. The developer notes that all changes addressed in version 4 are to be retained in the newest version of the registry and that "cardiac arrest data elements identified in the literature as risk factors will be included in the new version of the registry".

***Questions for the Committee**:*

- *How have the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How do you think participants will be impacted once changes have been completed to version 5 of the CathPCI registry?*

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

RATIONALE: N/A

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

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.
4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.
Improvement results No data on improvement was provided. The developer states that "this measure does not readily lend itself to improvement across the entire population of hospitals over time because evolutions in technology (e.g. circulatory support) enables sicker patients to be treated. However, the measure does have the opportunity to identify hospitals with higher mortality rates than expected".
4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).
Unexpected findings (positive or negative) during implementation [unexpected findings]
As noted in 4a2, the developer mentioned previous concerns from the interventional community that have been addressed in version 4 of the CathPCI Registry. The developer notes that all changes address in version 4 are to be retained the newest version of the registry and that "cardiac arrest data elements identified in the literature as risk factors will be included in the new version of the registry".

Potential harms The developer did not indicate any potential harms or benefits from this measure.
Additional Feedback: No additional feedback was provided.

***Questions for the Committee**:*

- *Do you think there are any benefits or harms from the continued implementation of this measure?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

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

RATIONALE: N/A

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

4a. Use:

- Feedback is publicly reported in two programs.
- The measure is used by Anthem
- As a decade old measure the feedback has declined with time
- It is being publicly reported as part of two programs and they are receiving feedback on the measure.
- considerable literature is present discussing the impact of public reporting of this measure. the developer uses the wording "safer PCI"" associated with lower mortality which may be the most accurate interpretation of this measure
- There is a notation about harmonization with measure 0536 30-day mort for PCI for STEMI or Cardiogenic Shock(has not occurred yet)
- Medium

4b. Usability:

- It is important to continue reporting this measure
- Benefits outweigh harms. Avoidance of intervention on very sick patients who could benefit is a potential harm, but this apparently has not happened.
- In the published study, there were no measured harms.
- This can be a usable metric for improvement. The only concern would be a higher risk demographic population that might skew the data for one facility versus another.
- while this measure does use a hard endpoint, there is no data to suggest that implementation of this measure has/will improve outcomes. few outliers likely drive most of the observed differences in mortality, yet this is where the measure has the greatest difficulty with the numerator/denominator/exclusion criteria/competing measures and potential for missing data/data loss
- There is a new version of the CathPCI registry that has not been tested for this measure. It supposedly addresses "previous concerns" from the interventional community (risk factors related to cardiac arrest)
- medium level

Criterion 5: Related and Competing Measures

Related or competing measures

Related measures:

- 0119 : Risk-Adjusted Operative Mortality for CABG
- 0230 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older
- 2411: Comprehensive Documentation for Indications for PCI

- 2459 In-hospital Risk Adjusted Rate of Bleeding Events for Patients Undergoing PCI
- 0535: 30-day All-Cause Risk-Standardized Mortality Rate Following PCI for Patients Without STEMI and Without Cardiogenic Shock
- 0536: 30-Day All-Cause Risk-Standardized Mortality Rate Following PCI for Patients with STEMI or Cardiogenic Shock

Harmonization

The developer notes that while the patient population is similar between this measure and measure #0119, the outcomes are slightly different and the method of revascularization is different, therefore harmonization is not possible. No other information was provided for the other measures with similar patient populations.

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

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 10, 2018

No comments have been submitted as of this date.

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

0133_NQF_evidence_attachment_20171108.pdf

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission? Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

1a Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0133

Measure Title: In-Hospital Risk Adjusted Rate of Mortality for Patients Undergoing PCI

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 11/8/2017

Instructions

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

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

1a. Evidence to Support the Measure Focus

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

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.

- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

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

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

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

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

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

Outcome

☒ Outcome: Risk Adjusted Mortality for PCI

☐ Patient-reported outcome (PRO):

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

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

☐ Process:

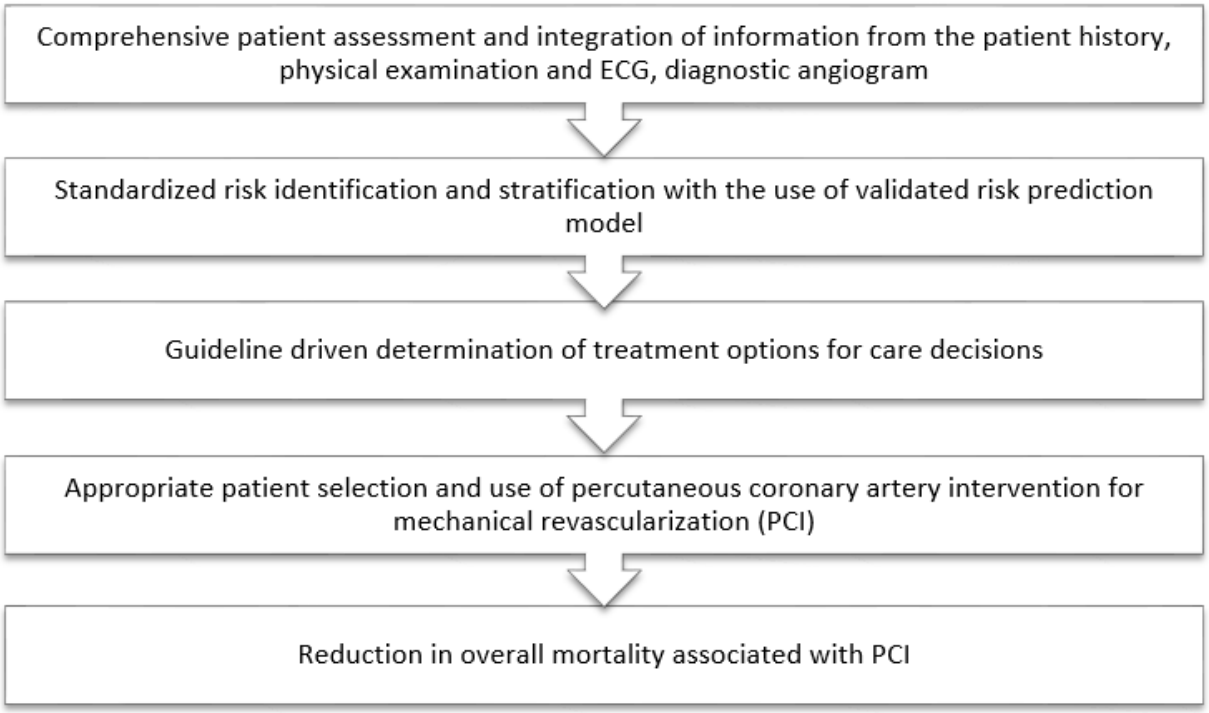
☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2. LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Comprehensive, personalized risk assessment can lead to decreased mortality in the PCI patient population.



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

Comprehensive patient assessment

There is a socio-economic demand that cardiologists provide services in an accountable and cost-effective fashion. There are opportunities to improve the health of individuals and of populations by better coordination of all aspects of medical care, and by assessing and responding to each individual's health risks. This goal of individualized care can be achieved for patients requiring coronary artery revascularization with the use of risk models that identify PCI related risk factors and accurately quantify the procedural risks. Demographic, pre-catheterization clinical variables, and angiographic variables are incorporated in the CathPCI Registry PCI Risk Adjusted mortality measure to obtain this personalized patient care [1].

Standardized risk identification and stratification with the use of validated risk prediction model

The NCDR CathPCI risk score was developed and validated from an analysis of data from 1,208,137 PCI procedures performed in the United States in a broad range of institutions from 2009 to 2011. Our initial risk prediction model included 21 variables, the eight with the strongest association with in-hospital mortality were included in the risk score: age, cardiogenic shock, prior heart failure, peripheral artery disease, chronic lung disease, glomerular filtration rate, NYHA functional class IV, and PCI status (STEMI or no STEMI). "This model indicates that In-hospital mortality was 1.4%, ranging from 0.2% among elective cases (45.1% of total cases) to 65.9% among patients with shock and recent cardiac arrest (0.2% of total cases). Cardiogenic shock and procedure urgency were the most predictive of inpatient mortality, whereas the presence of a chronic total occlusion, subacute stent thrombosis, and left main lesion location were significant angiographic predictors. The full, pre-catheterization, and bedside risk prediction models performed well in the overall validation sample (C-indexes 0.930, 0.928, 0.925, respectively) and among pre-specified patient subgroups" [2].

Several other risk models have been developed to predict in-hospital mortality and major complications after PCI. There are limitations within these older models. Some were validated within the institution at which they were developed creating concerns about generalizability. Other models were developed before the routine use of drug-eluting stents or aggressive antithrombotic therapy and based upon patient populations treated with balloon angioplasty alone. Finally

models focus on either acute coronary syndrome (ACS) or stable disease. These factors created limitations to the adoption of widespread use of these older models.

The National Heart, Lung and Blood Institute Dynamic Registry compared five multivariable mortality models for their ability to predict in-hospital mortality in 4448 patients undergoing PCI from 1997 to 1999. In this population, 64 patients died, for an in-hospital mortality rate of 1.4 percent. These models were developed and validated prior to the extensive use of stents in New York State, Northern New England Cooperative Group, Cleveland Clinic Foundation, and the University of Michigan as well as the original CathPCI Registry model [3]. The CathPCI Registry model has been updated and recalibrated. The volume of patients included in our model has expanded exponentially as it captures data from across the country.

The Mayo Clinic developed a risk score to identify patients at increased risk for major complications after PCI for procedures performed between January 1, 1996, and December 31, 1999. They validated their model in 2000. This model is relevant to current clinical practice since it was performed after stenting became routine, the patients were usually treated with clopidogrel or ticlopidine, and intravenous glycoprotein IIb/IIIa inhibitors were available. An updated report from the Mayo Clinic, using data from over 7457 PCI's performed between 2000 and 2005, developed two risk-prediction models, one for mortality alone and one for all major adverse cardiovascular events [4].

Risk scores for procedural death, defined as any death during the index hospitalization, and MACE contained the same 7 variables (age, myocardial infarction less than or equal to 24 hours, preprocedural shock, serum creatinine level, left ventricular ejection fraction, congestive heart failure, and peripheral artery disease). The two models successfully predicted the risk of adverse events during the index hospitalization. The model is useful for providing patients with individualized, evidence-based estimates of procedural risk as part of the informed consent process. However, one study limitation related to this model is that it includes performance at a single, referral center in a lower-risk patient population [4].

The EVENT registry evaluated 7592 consecutive patients who underwent successful or attempted PCI 47 hospitals throughout the United States between July 2004 and September 2006 to determine it pre-procedural cardiac troponin (cTn) elevation in patients with stable coronary artery disease was a predictor of adverse postprocedural outcome. The frequency of an elevated cTn immediately before PCI and its relationship to in-hospital and 1-year outcomes among patients who underwent PCI for either stable angina or a positive stress test was analyzed. The multivariable analyses adjusted for demographic, clinical, angiographic, and procedural factors, baseline cTn elevation was found to be independently associated with the composite of death or myocardial infarction at hospital discharge (odds ratio, 2.1; 95% confidence interval, 1.2 to 3.8; P=0.01) and at the 1-year follow-up (odds ratio, 2.0; 95% confidence interval, 1.2 to 3.3; P=0.005) [5]. This limitation to this model is a focus on either acute coronary syndrome (ACS) or stable disease.

Clinical acuity is a strong predictor of PCI procedural mortality. With inclusion of variables that further characterize clinical stability, the updated CathPCI Registry mortality models remains a current, and well-calibrated across the spectrum of PCI risk [2].

Guideline driven determination of treatment options for care decision

Guidelines described in sections 1a.3.

Appropriate patient selection and use of percutaneous coronary artery intervention for mechanical revascularization (PCI)

Appropriate use criteria (AUC) for coronary revascularization are tailored to the specific characteristics of individual patients. The evaluation of AUC covers broader array of specific conditions, sometimes hundreds for a given test or treatment decision, to encompass the majority of practice situations. Appropriateness relate to individual patient demographic characteristics, clinical history, risk scores, and/or symptoms and signs. "The increasing prevalence of coronary artery disease (CAD), continued advances in surgical and percutaneous techniques for revascularization and concomitant medical therapy for CAD, and the costs of revascularization have resulted in heightened interest regarding the appropriate use of coronary revascularization. Clinicians, payers, and patients are interested in the specific benefits

of revascularization. Inappropriate revascularization may be harmful to patients and generate unwarranted costs to the healthcare system, whereas appropriate revascularization procedures can improve patients' clinical outcomes" [7].

Reduction in overall mortality associated with PCI

Upon consideration of associated risk factors and evaluation of guidelines and appropriateness, coronary artery reperfusion improves clinical outcomes for patients. Statistics collected by the American Heart Association indicate that coronary heart disease caused an estimated 1 of every 6 deaths in the United States in 2008. Coronary heart disease mortality in 2008 was 405,309. Each year, an estimated 785,000 Americans will have a new MI, and approximately 470, 000 will have a recurrent MI. It is estimated that an additional 195 000 silent first MI occur each year. Approximately every 25 seconds, an American will have a coronary event, and approximately every minute, someone will die of one. The estimated direct and indirect cost of CVD for 2008 is $297.7 billion (MEPS, Agency for Healthcare Research and Quality, and NHLBI). In 2009, an estimated 7, 453,000 inpatient cardiovascular operations and procedures were performed in the United States (NHLBI tabulation of NHDS, NCHS). In-hospital death rates for PCI have remained stable although comorbidities increased for patients who received the procedure increases over time [8].

References

[1] Peterson ED, Dai D, DeLong ER, et al. Contemporary Mortality Risk Prediction for Percutaneous Coronary Intervention: Results From 588,398 Procedures in the National Cardiovascular Data Registry. J Am Coll Cardiol. 2010;55(18):1923-1932. doi:10.1016/j.jacc.2010.02.005.

[2] Brennan J, Curtis JP, Dai D, et al. Enhanced Mortality Risk Prediction With a Focus on High-Risk Percutaneous Coronary Intervention: Results From 1,208,137 Procedures in the NCDR (National Cardiovascular Data Registry). J Am Coll Cardiol Intv. 2013;6(8):790-799. doi:10.1016/j.jcin.2013.03.020.

[3] Modeling and risk prediction in the current era of interventional cardiology: a report from the National Heart, Lung, and Blood Institute Dynamic Registry. Holmes DR, Selzer F, Johnston JM, Kelsey SF, Holubkov R, Cohen HA, Williams DO, Detre KM, National Heart, Lung, and Blood Institute Dynamic Registry. Circulation. 2003;107(14):1871.

[4] Bedside estimation of risk from percutaneous coronary intervention: the new Mayo Clinic risk scores. Singh M, Rihal CS, Lennon RJ, Spertus J, Rumsfeld JS, Holmes DR Jr. Mayo Clin Proc. 2007;82(6):701.

[5] Prevalence and prognostic significance of preprocedural cardiac troponin elevation among patients with stable coronary artery disease undergoing percutaneous coronary intervention: results from the evaluation of drug eluting stents and ischemic events registry. Jeremias A, Kleiman NS, Nassif D, Hsieh WH, Pencina M, Maresh K, Parikh M, Cutlip DE, Waksman R, Goldberg S, Berger PB, Cohen DJ, Evaluation of Drug Eluting Stents and Ischemic Events (EVENT) Registry Investigators. Circulation. 2008;118(6):632.

[6] 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. *J Am Coll Cardiol.* 2011;58(24):e44-e122. doi:10.1016/j.jacc.2011.08.007.

[7] Patel MR, Dehmer GJ, Hirshfeld JW, Smith PK, Spertus JA. ACCF/SCAI/STS/AATS/AHA/ASNC/HFSA/SCCT 2012 Appropriate Use Criteria for Coronary Revascularization Focused Update: A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, Society for Cardiovascular Angiography and Interventions, Society of Thoracic Surgeons, American Association for Thoracic Surgery, American Heart Association, American Society of Nuclear Cardiology, and the Society of Cardiovascular Computed Tomography. *J Am Coll Cardiol.* 2012;59(9):857-881. doi:10.1016/j.jacc.2011.12.001.

[8] AHA Statistical Update: Heart Disease and Stroke Statistics—2012 Update: A Report From the American Heart AssociationCirculation. 2012; 125: e2-e220 Published online before print December 15, 2011, doi: 10.1161/ CIR.0b013e31823ac046: Retrieved at http://circ.ahajournals.org/content/125/1/e2.full#ref-288

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

N/A

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	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. *J Am Coll Cardiol.* 2011;58(24):e44-e122. doi:10.1016/j.jacc.2011.08.007. Available at: http://content.onlinejacc.org/article.aspx?articleid=1147816
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	5.2.1 UA/NSTEMI: Recommendations CLASS I: 3. The selection of PCI or CABG as the means of revascularization in the patient with ACS should generally be based on the same considerations as those without ACS . (Level of Evidence: B) 5.2.2.2 Primary PCI of the Infarct Artery: Recommendations CLASS I: 4. Primary PCI should be performed in patients with STEMI who develop severe heart failure or cardiogenic shock and are suitable candidates for revascularization as soon as possible, irrespective of time delay. (Level of Evidence: B) CLASS IIb: 1. Primary PCI might be considered in asymptomatic patients with STEMI and higher risk presenting between 12 and 24 hours after symptom onset. (Level of Evidence: C) Revascularization to Improve Survival Recommendations CLASS IIa: 1. PCI to improve survival is reasonable as an alternative to CABG in selected stable patients with significant (≥50% diameter stenosis) unprotected left main CAD with: 1) anatomic conditions associated with a low risk of PCI procedural complications and a high likelihood of good long-term outcome (e.g., a low SYNTAX score [≤22], ostial or trunk left main CAD); and 2) clinical characteristics that predict a significantly increased risk of adverse surgical outcomes (e.g., STS-predicted risk of operative mortality ≥5%) . (Level of Evidence: B) 2. PCI to improve survival may be reasonable as an alternative to CABG in selected stable patients with significant (≥50% diameter stenosis) unprotected left main CAD with: 1) anatomic conditions associated with a low to intermediate risk of PCI procedural complications and an intermediate to high likelihood of good long-term outcome (e.g., low-intermediate SYNTAX score of <33, bifurcation left main CAD); and 2) clinical characteristics that predict an increased risk of adverse surgical outcomes (e.g., moderate-severe chronic obstructive pulmonary disease, disability from previous stroke, or previous cardiac surgery; STS-predicted risk of operative mortality >2%). (Level of Evidence: B)

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Recommendation classification is listed with each of the 5 recommendations above for clarity. ACCF/AHA/SCAI recommendations have been assigned a Class I and Class II 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." Class II recommendations refer to "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".
Provide all other grades and definitions from the evidence grading system	• 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 **(See Table 1 below)**
Grade assigned to the **recommendation** with definition of the grade	The section of the ACCF/AHA/SCAI PCI guideline which includes the recommendations referenced above pertains to the necessity to evaluate a patients risk factors and incorporating ACS symptoms and hemodynamic stability into the determination for appropriate revascularization The weight of the evidence in support of the ACCF/AHA/SCAI recommendations is rated as Level B and Level C as noted parenthetically. Level B evidence refers to "Data derived from a single randomized trial, or nonrandomized studies" while Level C evidence refers to "Only consensus opinion of experts, case studies, or standard-of-care."
Provide all other grades and definitions from the recommendation grading system	**(See Table 1 below)**
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Observational hypothesis-generating analysis, pooled meta- analyses of randomized trials, multi- centered, and single centered study designs were all deployed in the creation of these guidelines. Information regarding the overall quality of evidence across the studies does not exist.
Estimates of benefit and consistency across studies	The focus of the data included for these guidelines were the benefits of treating the Left Main via PCI and the benefits of DES vs. BMS for device choice. The guidelines included here demonstrate the importance of evaluating the personal risk of for PCI. The process of the PCI is not the intent of this measure but only one step in the decision making process of achieving a decreased mortality for this patient population. Thus an extensive evaluation of the magnitude and direction of effect within the evidence used to support PCI was not conducted for the purpose of this application.
What harms were identified?	An extensive evaluation of the harms described within the evidence used to support PCI was not conducted for the purpose of this application.

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	The ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention has not been updated since the 2011 document referenced in the citations above.

Table 1

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 a 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 considered may/might be reasonable usefulness/effectiveness is unknown/unclear/ uncertain or not well established	COR III: No Benefit — COR III: Harm is not recommended — potentially harmful is not indicated — causes harm should not be performed/ administered/ other — associated with excess morbidity/ mortality is not useful/ beneficial/ effective — 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		

1a.4 OTHER SOURCE OF EVIDENCE

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

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

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

This measure allows benchmarking against the national aggregate and against hospitals with similar volume, so that hospitals with high mortality rates can engage in quality improvement to reduce mortality following PCI procedures. In-depth analysis of the causes behind variations in mortality during or post PCI can lead to the identification of best practices. Particularly actionable opportunities to improve care are to reduce peri-procedural bleeding rates and acute kidney injury, where operators have the option to pursue strategies that decrease these complications. In addition, detailed case reviews can identify operators with poorer performance for whom additional training or reduced caseloads could be considered. Active dissemination of those best practices and support to enable their adoption will improve outcomes and reduce variations in clinical practice. Improvements in the quality of care resulting from the evaluation of the risk for mortality, before and after implementing quality improvement interventions, can enable centers to quantify their improved outcomes with respect to peri-procedural mortality and a reduction in cost associated with these events. Additionally, by putting the responsibility for improved quality in the hands of physicians and other health-care practitioners, this risk-adjusted mortality measure engages the medical community around the common goal of better health-care value.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

2015 Data:

Data range date: Quarter 1 through Quarter 4, 2015

Number of patients: 699,049

Number of PCI procedures per hospital volume:

0-10: 5 hospitals; 35 Patients

11-200: 405 hospitals; 45826 Patients

201-400: 456 hospitals; 134174 Patients

401-600: 274 hospitals; 134305 Patients

601-1000: 260 hospitals; 197675 Patients

1001-2000: 138 hospitals; 180023 Patients

2001+: 15 hospitals; 35607 Patients

Deciles:

Mean: 1.95%

Stddev: 1.07%

Quartile 1: 1.33%

Quartile 3: 2.32%

Deciles of mortality adjusted rates:

1: 0.96%

2: 1.23%

3: 1.42%

4:1.59%

5(median): 1.76%

6: 1.95%

7: 2.17 %

8: 2.47%

9: 3.0%

In 2015, we observed a >3-fold variation in mortality rates from the lowest to highest decile; a significant opportunity to improve survival by an absolute percentage of 2.04% in the highest decile.

2016 Data:

Data range date: Quarter 1 through 4 2016

Number of patients: 722,029

Number of PCI procedures per hospital volume:

0-10: 10 hospitals; 34 Patients

11-200: 438 hospitals; 48342 Patients

201-400: 453 hospitals; 132373 Patients

401-600: 280 hospitals; 137916 Patients

601-1000: 278 hospitals; 210683 Patients

1001-2000: 147 hospitals; 193893 Patients

2001+: 12 hospitals; 29357 Patients

Adjusted Rate statistics:

Mean: 1.9%

Stddev: 1.06%

Quartile 1: 1.30%

Quartile 3: 1.76%

Deciles of mortality adjusted rates:

1: 0.92%

2: 1.2%

3: 1.41%

4:1.58%

5(median):1.76%

6: 1.95%

7: 2.15%

8: 2.45%

9: 2.96%

In 2016, we observed a >3-fold variation in mortality rates from the lowest to highest decile; a significant opportunity to improve survival by an absolute percentage of 2.04% in the highest decile. Collectively, these 2 years of data, supplemented with prior evidence of declining mortality rates with PCI, show that there have been progressive improvements in the safety of PCI over time. However, there have not been formal evaluations of strategies to move poorer performing sites to better performance. This is a demonstration for an opportunity for improvement based on the noted performance scores, but further efforts to decrease other complications of PCI, such as bleeding and acute kidney injury, are opportunities for further improving performance across the US.

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

N/A

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

While we observed some statistically significant differences by gender, race and insurance status, the absolute rates after patient-level adjustment for mortality risk were modest. Of particular interest is that when compared with the expected mortality rates, those with private insurance had significantly better survival, while those with all other insurance types did worse. Similarly, suburban and rural hospitals seemed to provide safer PCI than urban centers. The difference by race and gender between observed and predicted rates were very small.

Disparities Data

The information below provides the observed vs. predicted rates of mortality for various populations that include hospital location, sex, insurance status, and race.

Data range date: Quarter 1 through 4, 2016

	Total	Hospital Location			
	n = 722029	RURAL n = 101625	SUBURBAN n = 228504	URBAN n = 391900	P-Value
Mortality					
Observed Mortality	13406 (1.8567%)	1800 (1.7712%)	4071 (1.7816%)	7535 (1.9227%)	< 0.001
Estimated Probability	1.831 ± 0.071%	1.8 ± 0.069%	1.8 ± 0.069%	1.87 ± 0.07%	< 0.001
Continuous variables compared using one-way analysis of variance. Categorical variables compared using chi-square or Fisher's exact test.					

Hospital Location	OE Ratio
RURAL	1.00031
SUBURBAN	0.99021
URBAN	1.03039

	Total	Teaching Hospital			
	n = 722029	1 n = 345237	0 n = 376792	P-Value	
Mortality					
Observed Mortality	13406 (1.8567%)	6651 (1.9265%)	6755 (1.7928%)	< 0.001	
Estimated Probability	1.83 ± 0.07%	1.87 ± 0.07%	1.79 ± 0.07%	< 0.001	
Continuous variables compared using Student's T-test. Categorical variables compared using chi-square or Fisher's exact test.					

Teaching Hospital	OE Ratio
Non-teaching	1.00033
Teaching	1.02787

	Total	Sex			
	n = 722029	Male n = 496990	Female n = 225039	P-Value	
Mortality					
Observed Mortality	13406 (1.8567%)	8244 (1.6588%)	5162 (2.2938%)	< 0.001	
Estimated Probability	1.83 ± 0.07%	1.69 ± 0.07%	2.15 ± 0.08%	< 0.001	
Continuous variables compared using Student's T-test. Categorical variables compared using chi-square or Fisher's exact test.					

Sex	OE Ratio
(1) Male	0.98175
(2) Female	1.06958

	Total	**inscat**					
	n = 722029	**1 Private n = 470179**	**2 Medicare n = 165245**	**3 Medicaid n = 37896**	**4 Other n = 16076**	**5 None n = 32633**	**P-Value**
Mortality							
Observed Mortality	13406 (1.8567%)	7465 (1.5877%)	4094 (2.4775%)	616 (1.6255%)	282 (1.7542%)	949 (2.9081%)	< 0.001
Estimated Probability	1.83 ± 0.07%	1.64 ± 0.066%	2.37 ± 0.08%	1.55 ± 0.06%	1.75 ± 0.07%	2.24 ± 0.08%	< 0.001
Continuous variables compared using one-way analysis of variance. Categorical variables compared using chi-square or Fisher's exact test.							

inscat	OE Ratio
1 Private	0.96922
2 Medicare	1.04408
3 Medicaid	1.04802
4 Other	1.00064
5 None	1.29890

	Total	racecat			
	n = 722029	1 Caucasian n = 621359	2 Af Am n = 62268	3 Other n = 38402	P-Value
Mortality					
Observed Mortality	13406 (1.8567%)	11457 (1.8439%)	1153 (1.8517%)	796 (2.0728%)	0.005
Estimated Probability	1.83 ± 0.07%	1.82 ± 0.07%	1.82 ± 0.07%	2.1 ± 0.08%	< 0.001
Continuous variables compared using one-way analysis of variance. Categorical variables compared using chi-square or Fisher's exact test.					

racecat	OE Ratio
1 Caucasian	1.01462
2 Af Am	1.01788
3 Other	0.99648

While we observed some statistically significant differences by hospital location, gender, race and insurance status, the absolute rates after patient-level adjustment for mortality risk were small. Of particular interest is that when compared with the expected mortality rates, those with private insurance had slightly better survival (4% better than expected, while those with Medicare and Medicaid did slightly worse (5% worse) and those without insurance did substantially worse (30% worse than expected). Similarly, suburban and rural hospitals seemed to provide safer PCI than urban centers. The difference by race and gender between observed and predicted rates were very small.

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

Finding minimal differences by patient characteristics, as compared with differences across deciles of hospital performance further supports the idea that hospital-focused quality improvement efforts, rather than patient-specific ones, are likely to have the greatest impact on improving the quality and safety of PCI. These findings also do not support the need to stratify the 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 sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

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

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

Cardiovascular, Cardiovascular : Coronary Artery Disease (PCI)

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

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

Populations at Risk

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

https://www.ncdr.com/webncdr/cathpci/home/datacollection:ACC realizes the various NCDR endorsed measures are not readily available on their own main webpage. However, ACCF plans to update their main webpage (cardiosource.org)

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:** cathpci_v4_codersdictionary_4-4.pdf

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

No changes were made to the measure specification since the 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) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Patients 18 years of age and older with a PCI procedure performed during episode of care who expired

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

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

PCI=yes

Coding instructions to identify patients in the numerator: indicate if the patient had a percutaneous coronary intervention (PCI) Selection options: yes/no

Supporting definitions: PCI: A percutaneous coronary intervention (PCI) is the placement of an angioplasty guide wire, balloon, or other device (e.g. stent, atherectomy, brachytherapy, or thrombectomy catheter) into a native coronary artery or coronary bypass graft for the purpose of mechanical coronary revascularization. Source: NCDR

AND

Discharge status=deceased

Response options: Alive/deceased

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

Patients 18 years of age and older with a PCI procedure performed during episode of care.

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

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

PCI=yes

Coding instructions for identifying the measure's denominator: indicate if the patient had a percutaneous coronary intervention (PCI); Selection options: yes/no

Supporting definitions: PCI: A percutaneous coronary intervention (PCI) is the placement of an angioplasty guide wire, balloon, or other device (e.g. stent, atherectomy, brachytherapy, or thrombectomy catheter) into a native coronary artery or coronary bypass graft for the purpose of mechanical coronary revascularization. Source: NCDR

AND

Age>=18: patients must be 18 years of age to be included in the registry.

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

1. NCDR Registry patients who did not have a PCI (Patient admissions with a diagnostic cath only during that admission);

2. Patient admissions with PCI who transferred to another facility on discharge

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

See coding instructions for defining a PCI in S.5, which also apply to the denominator. In addition, it is important to note that all data submissions must pass the data quality and completeness reports to be included. Note: If one or two variables are missing, the value is imputed for certain characteristics . In our data quality program, all key variables in the risk model have a high "inclusion" criteria. This means that, when a hospital submits data to us , they need to have a high level of completeness (around 95-99%) for those variables. If they are not able to meet the criteria in our data quality program, they do not receive risk-adjusted mortality for any of the records they submitted for that quarter.

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

N/A: We do not use univariate categorizations to apply the measure to subsets of the population. Rather, we use a statistical risk model to integrate all patient characteristics prior to calculating the outcome.

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

Other

If other: We have used hierarchical logistic regression to calculate the risks for peri-procedural mortality and use these data to create risk-standardized event rates.

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

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

1. Remove hospitals who fail data quality and completeness reports as outlined in the NCDR Data Quality Program (further discussed in the Testing Supplement and described in section S.9 above)

2. Count of admissions from data submissions that pass NCDR data inclusion thresholds.

3. Remove patient's subsequent PCIs during the same admission (if the patient had more than one PCI procedure during that admission). (Note: The measure consists of the first PCI in a hospital stay and subsequent PCI in that stay are not included in the denominator)

4. Remove admissions without PCI during admission

5. Remove patient admissions with PCI who transferred to another facility on discharge;

6. Calculate measure using weight system based on predictive variables as outlined in the accompanying testing documents and supplemental materials.

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Registry Data

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

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

National Cardiovascular Data Registry Percutaneous Coronary Interventions

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

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

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

Facility

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

Inpatient/Hospital

If other:

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

2. Validity – See attached Measure Testing Submission Form

0133_NQF_testing_attachment_20171108.pdf

2.1 <u>For maintenance of endorsement</u>

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

No

2.2 <u>For maintenance of endorsement</u>

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

No

2.3 <u>For maintenance of endorsement</u>

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

Yes - Updated information is included

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

Measure Number (*if previously endorsed*)**:** 0133
Measure Title: In-Hospital Risk Adjusted Rate of Mortality for Patients Undergoing PCI
Date of Submission: <u>11/8/2017</u>

Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.

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

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

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

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

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

AND

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

Notes

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

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

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

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

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

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

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

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

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

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

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

We propose to use a clinical registry, the National Cardiovascular Data Registry (NCDR) for CathPCI Registry. This is a national quality improvement registry that is currently participated in >1,300 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.

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

Since this model has already been approved by NQF as a performance measure, we have performed additional testing, in new data, to establish its continued value and accuracy as a performance measure.

We have chosen to use different datasets to provide support for different aspects of the proposed measure.

1. Audit data: 01/2009-12/ 2009 has been previously used to support the inter-rater reliability of the application. It was established that there is high inter-rater reliability, as compared with independent chart audits and found >90% accuracy for most variables. Please see prior submission for these data.

2. Creation of the Mortality model was performed on all national NCDR data from 07/2009–06/2011 and has been used to provide a description and initial performance characteristics of the model.

3. A separate cohort of the NCDR CathPCI registry was used to validate the model, which included all data collected during the 2012 calendar year (01/2012-12/2012). These data were also used to provide test-retest reliability of the data elements for the risk model and further validation of the relationship between the predictor variables and mortality, including additional data supporting the discrimination and calibration of the model.

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

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

Creation of the Mortality Derivation and Validation model:

1,253 hospitals were included. See additional information under section 1.6.

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*)

For the updated derivation and validation of the mortality risk model, 1,208,137 patients undergoing PCI between 7/2009-06/2011 at 1,253 hospitals were included; 60% in the derivation cohort and a random 40% in the validation cohort. In-hospital mortality was 1.4%, ranging from 0.2% among elective cases (45.1% of total cases) to 65.9% among

patients with shock and recent cardiac arrest (0.2% of total cases) . A summary of these patients' clinical characteristics and the hospital characteristics are provided under Table 1 and 2.

Table 1. Derivation and Validation Characteristics

	Development (N=724,883)	Validation (N=483,254)
Overall population	1.38	1.40
MI status		
STEMI	5.22	5.33
No STEMI	0.65	0.65
Gender		
Men	1.22	1.23
Women	1.72	1.74
Age group		
Age > 70 yrs	2.23	2.25
Age ≤ 70 yrs	0.96	0.98
Diabetes status		
Diabetes mellitus	1.51	1.50
No diabetes mellitus	1.31	1.34
Cardiac arrest	24.32	25.07
Cardiogenic shock and PCI status		
Sustained shock and salvage	63.99	68.85
Sustained shock or salvage	33.45	34.33
Transient shock but not salvage	15.26	14.85
Emergency PCI without shock/salvage	2.26	2.29
Urgent PCI without shock/salvage	0.63	0.63
Elective PCI without shock/salvage	0.18	0.18
STEMI = ST-segment elevation myocardial infarction; All other abbreviations can be found in Table 1.		

Table 2. Hospital Characteristics

2012 Data

	Total n = 1367
Participant Classification	
FREE STANDING CATH LAB	1 (0.1%)
FREE STANDING CATH LAB/CLINIC	3 (0.2%)
HEALTH SYSTEM/NETWORK	60 (4.4%)
HOSPITAL	1203 (88.1%)
HOSPITAL/HEALTH NETWORK	95 (7.0%)
OTHER	3 (0.2%)
PRIVATE CV PRACTICE	1 (0.1%)
Missing	1
Hospital Location	
RURAL	249 (18.2%)
SUBURBAN	492 (36.0%)
URBAN	625 (45.8%)
Missing	1
Participant Type	
GOVERNMENT	21 (1.5%)
PRIVATE/COMMUNITY	1232 (90.2%)
UNIVERSITY	113 (8.3%)
Missing	1
Teaching Hospital	524 (38.4%)
Missing (.)	1
Public Hospital	530 (38.8%)
Missing (.)	1
Volume (Med (IQR))	367 (188, 643)
Census Region	
MIDWEST REGION	395 (28.9%)
NORTHEAST REGION	182 (13.3%)
SOUTH REGION	521 (38.2%)
WEST REGION	267 (19.6%)
Missing	2

For the additional testing of predictive validity, calibration and test-retest reliability, we used 634,084 patients undergoing PCI between 1/2012-12/2012, of whom 10,212 (1.6%) had a mortality event. A summary of these patients' clinical characteristics (focusing upon those that are predictor variables in the final, full model) are provided under Table 3.

Table 3. Predicted Probability of Mortality

	Total	Observed Mortality		
	n = 634084	Yes n = 10212	No n = 623872	P-Value
Mortality				
Predicted Probability of Death	0.01583 ± 0.06557	0.26718 ± 0.27434	0.01171 ± 0.04569	< 0.001
Mortality Variables				
STEMI	111775 (17.6%)	6275 (61.4%)	105500 (16.9%)	< 0.001
Age	64.8 ± 12.1	70.4 ± 12.8	64.7 ± 12.0	< 0.001
Body Mass Index	30.0 ± 6.4	28.5 ± 7.0	30.0 ± 6.4	< 0.001
CVD	79750 (12.6%)	1771 (17.3%)	77979 (12.5%)	< 0.001
PVD	79224 (12.5%)	1800 (17.6%)	77424 (12.4%)	< 0.001
Prior PCI	258993 (40.8%)	2665 (26.1%)	256328 (41.1%)	< 0.001
Left Ventricular Ejection Fraction	52.6 ± 10.1	43.2 ± 13.6	52.8 ± 10.0	< 0.001
GFR	71.4 ± 18.0	55.9 ± 19.9	71.7 ± 17.9	< 0.001
Non-insulin Diabetes vs. No Diabetes	145444 (22.9%)	2256 (22.1%)	143188 (23.0%)	0.040
Insulin Diabetes vs. No Diabetes	91052 (14.4%)	1811 (17.7%)	89241 (14.3%)	< 0.001
HF NYHA Class I/II/III w/in 2 Wks vs. No HF	49822 (7.9%)	1250 (12.2%)	48572 (7.8%)	< 0.001
HF NYHA Class IV w/in 2 Wks vs. No HF	16595 (2.6%)	1889 (18.5%)	14706 (2.4%)	< 0.001
Cardiac arrest w/in 24 hrs	13637 (2.2%)	3559 (34.9%)	10078 (1.6%)	< 0.001
Salvage Status and Carshock w/in 24hrs and at start of PCI vs. Elective Status and No Carshock	1563 (0.2%)	1072 (10.5%)	491 (0.1%)	< 0.001
Salvage Status or Carshock w/in 24hrs and at start of PCI (not both) vs. Elective Status and No Carshock	9115 (1.4%)	3159 (30.9%)	5956 (1.0%)	< 0.001
Carshock w/in 24hrs or at start of PCI (not both) vs. Elective Status and No Carshock	8460 (1.3%)	1295 (12.7%)	7165 (1.1%)	< 0.001
Emergent Status and No Carshock vs. Elective Status and No Carshock	107052 (16.9%)	2440 (23.9%)	104612 (16.8%)	< 0.001
Urgent Status and No Carshock vs. Elective Status and No Carshock	254411 (40.1%)	1757 (17.2%)	252654 (40.5%)	< 0.001
pLAD vs. Other	107452 (16.9%)	2732 (26.8%)	104720 (16.8%)	< 0.001
Left Main vs. Other	14756 (2.3%)	831 (8.1%)	13925 (2.2%)	< 0.001
In-stent Thrombosis on some lesion previously treated w/in 1 month	2044 (0.3%)	152 (1.5%)	1892 (0.3%)	< 0.001
Number of Diseased Vessels (2,3) vs. (0,1)	260307 (41.1%)	6245 (61.2%)	254062 (40.7%)	< 0.001
Chronic Total Occlusion	19138 (3.0%)	597 (5.8%)	18541 (3.0%)	< 0.001
History				
Intra-Aortic Balloon Pump	15357 (2.4%)	3864 (37.8%)	11493 (1.8%)	< 0.001
Missing (.)	169	1	168	

	Total	Observed Mortality		
	n = 634084	Yes n = 10212	No n = 623872	P-Value
Prior MI	193054 (30.5%)	2808 (27.6%)	190246 (30.5%)	< 0.001
Missing (.)	173	21	152	
Prior PCI	258993 (40.9%)	2665 (26.1%)	256328 (41.1%)	< 0.001
Missing (.)	150	13	137	
Currently on Dialysis	15882 (2.5%)	665 (6.5%)	15217 (2.4%)	< 0.001
Missing (.)	605	21	584	
Chronic Lung Disease	97244 (15.3%)	2202 (21.6%)	95042 (15.2%)	< 0.001
Missing (.)	322	23	299	
Diabetes Mellitus	236496 (37.3%)	4067 (39.9%)	232429 (37.3%)	< 0.001
Missing (.)	297	16	281	
Cath Lab Visit				
PCI Indication				
Immediate PCI for STEMI	97691 (15.4%)	5516 (54.0%)	92175 (14.8%)	< 0.001
PCI for STEMI (Unstable, >12 hrs from Sx onset)	5944 (0.9%)	498 (4.9%)	5446 (0.9%)	
PCI for STEMI (Stable, >12 hrs from Sx onset)	2546 (0.4%)	72 (0.7%)	2474 (0.4%)	
PCI for STEMI (Stable after successful full-dose Thrombolysis)	2221 (0.4%)	15 (0.1%)	2206 (0.4%)	
Rescue PCI for STEMI (after failed full-dose lytics)	3364 (0.5%)	172 (1.7%)	3192 (0.5%)	
PCI for high risk Non-STEMI or unstable angina	332909 (52.5%)	3231 (31.6%)	329678 (52.9%)	
Staged PCI	34929 (5.5%)	98 (1.0%)	34831 (5.6%)	
Other	154318 (24.3%)	607 (5.9%)	153711 (24.6%)	
Missing (.)	162	3	159	
CAD Presentation				
No symptom, no angina	35865 (5.7%)	328 (3.2%)	35537 (5.7%)	< 0.001
Symptom unlikely to be ischemic	14307 (2.3%)	119 (1.2%)	14188 (2.3%)	
Stable angina	89810 (14.2%)	137 (1.3%)	89673 (14.4%)	
Unstable angina	249827 (39.4%)	934 (9.1%)	248893 (39.9%)	
Non-STEMI	134840 (21.3%)	2539 (24.9%)	132301 (21.2%)	
ST-Elevation MI (STEMI) or equivalent	109286 (17.2%)	6152 (60.3%)	103134 (16.5%)	
Missing (.)	149	3	146	
Heart Failure w/in 2 Weeks	66417 (10.5%)	3139 (30.8%)	63278 (10.1%)	< 0.001
Missing (.)	275	10	265	
Cardiomyopathy or Left Ventricular Systolic Dysfunction	68481 (10.8%)	2274 (22.3%)	66207 (10.6%)	< 0.001
Missing (.)	156	2	154	
Pre-operative Evaluation Before Non-Cardiac Surgery	11823 (1.9%)	101 (1.0%)	11722 (1.9%)	< 0.001
Missing (.)	223	2	221	

	Total	Observed Mortality		
	n = 634084	Yes n = 10212	No n = 623872	P-Value
Cardiogenic Shock w/in 24 Hours	13197 (2.1%)	4528 (44.3%)	8669 (1.4%)	< 0.001
Missing (.)	109		109	
Cardiac Arrest w/in 24 Hours	13637 (2.2%)	3559 (34.9%)	10078 (1.6%)	< 0.001
Missing (.)	178	3	175	
Pre-PCI Left Ventricular Ejection Fraction	52.4 ± 12.5	37.8 ± 16.1	52.5 ± 12.4	< 0.001
Missing	188804	5396	183408	
Outcomes				
Discharge Status				
Alive	623872 (98.4%)	0 (0.0%)	623872 (100.0%)	< 0.001
Deceased	10212 (1.6%)	10212 (100.0%)	0 (0.0%)	

Continuous variables compared using Student's T-test.

Categorical variables compared using chi-square or Fisher's exact test.

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 the same data described above for all aspects of this supplement, except for the test-retest reliability of the data elements, where we restricted the sample to those with 2 procedures in 2012.

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

Social risk factors were not used in this risk model for the following reasons. First, as a detailed clincial registry used for quality assessment and improvement, there are not prospective interviews with patients to obtain patient-reported data. Second, while proxy variables could be considered, these were not felt to be relevant to an inpatient mortality model, in contrast to a longer-term outcome model where difficulties with access to care, affording medications or cardiac rehabilitation would be more important. Moreover, while it may be true that worse social risk factors might be associated with more severe illness at the time of presentation, we had direct access to detailed clinical variables describing the severity of illness and feel that incorporating such factors (e.g. cardiogenic shock, cardiac arrest, etc.) is a much more accurate means of stratifying risk. Accordingly, we feel that in this model of in-hospital mortality, given the rich clinical data available through the NCDR CathPCI registry, that social risk factors, which are not readily available, would not likely contribute much improvement to this particular risk model.

2a2. RELIABILITY TESTING

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

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

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

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

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

Measure Score:

ACCF performed the signal-to-noise analysis on the same cohort of individuals as noted under Section 1.3. (testing method 3). Only hospitals with a minimum of 10 eligible patients were included in the analysis to prevent undetected bias introduced by the inclusion of hospitals with a small sample size.

Data Element:

ACCF evaluated the test-retest reliability by reviewing CathPCI patients who were readmitted or had a repeat procedure in 2012. 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.

Data Element:

The NCDR Data Quality Program ensures that data submitted to the NCDR are complete validly collected. 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 4.

Table 4. 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 Data Quality Report 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. 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*)

Signal to Noise Analysis:

Signal to Noise analysis for the hospitals are noted under Table 5.

Table 5. Signal to Noise Analysis

Level	Signal-to-Noise
All, >10 Procedures	.537
>Q1 (>181 Procedures)	.582
>Q2 (>356 Procedures)	.659
>Q3 (>626 Procedures)	.748
>Average (>467 Procedures)	.700

Assessment of test-retest reliability among patients undergoing 2 procedures within 2012:

The key data elements for the mortality risk model tested among patients with 2 procedures in 2012 are shown below:

Gender demonstrated excellent reproducibility, with only 12 of 40,197 (0.03%) patients having different genders on the 2 procedures.

Age as assessed by Date of Birth was identical in 99.91% of the 40,045 patients on both assessments.

Cerebrovascular disease (CVD) revealed that only 1160 patients had evidence of CVD on the initial visit that was not noted on the second visit. This represents 2.9% of the population being clearly misclassified on one of the assessments.

Peripheral Vascular Disease (PVD) revealed that only 1332 (3.3%) patients who had evidence of PVD at the time of their initial PCI no longer had this recorded at the time of their second procedure and were clearly misclassified on one of the assessments.

Chronic Lung Disease (CLD) was recorded in 1366 (3.4%) of the patients at the time of their initial PCI, but not at the time of the second procedure.

Prior PCI should have been recorded on the second procedure for each of the 40,045 patients. 987 (2.5%) were not classified as having had a prior PCI.

Diabetes was not recorded among 731 (1.8%) of the patients who were noted to have diabetes at the time of their original procedure.

Because dynamic elements are expected to change over time, the following variables could not have their test-retest reliability assessed by this method: Prior cardiac arrest, GFR, NYHA classification, shock within 24 hours of PCI, indication

for PCI (e.g. STEMI vs. NSTEMI vs. others), urgency of the procedure, number, appearance and location of diseased vessels, lesion severity as assessed by the SCAI definitions, BMI, and TIMI flow could not be assessed using this approach.

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 Analysis:

The signal to noise ratio analysis measures the confidence levels in differentiating performance between hospitals. These numbers demonstrate variability that is attributable to real differences in hospital quality as opposed to measurement error.

Assessment of test-retest reliability among patients undergoing 2 procedures within 2012:

Finding no clear misclassification by test-retest reliability for any assessable risk factor being >3.5% provides strong support for the test-retest reliability of the mortality risk factors assessed.

Collectively, we believe that the prior audit data and repeat procedure data strongly support the reliability of the data elements used in the model.

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

_____

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

Rationale for proposing this outcome - Peri-procedural mortality is the most dreaded complication of PCI. The currently NQF-approved risk-adjusted peri-procedural mortality model has excellent discrimination and markedly shifts the observed performance of hospitals (see Section 2b4.9 below) by accounting for patient characteristics present prior to the conduct of the procedure. Given the marked distribution of performance across hospitals, we believe that some hospitals are clearly performing PCI more safely than others and that there is great opportunity to improve the safety of PCI at some centers. Importantly, we have also created a much simpler, pre-procedural risk model that can be used clinically to assess patients' risks for mortality. These estimates can be used by heart teams to define the best treatment strategy for each patient.

Content validity of this outcome –the specific definition used in defining peri-procedural mortality and the inclusion/exclusion criteria were was achieved by the specialized expertise of those individuals who developed this model as well as the structured discussions that the group conducted(Peterson et al. *J Am Coll Cardiol* 2010; 55: 1923-32). For this particular topic those individuals who were involved in identifying the key attributes and variables for this risk model were leaders and experts in the field of interventional cardiology. Serial phone calls were held to be both define the event and to examine and vet the risk model. Additional review was provided by the following 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 mortality as an 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.

Dr. Jeptha Curtis (chair), Dr. Frederick Masoudi, Dr. John Rumsfeld, Dr. David Malenka, and Dr. Issam Moussa.

NCDR Registry Steering 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.

Dr. Issam D. Moussa (chair), Dr. Kirk N. Garratt, Dr. Lloyd W. Klein, Dr. Kendrick A. Shunk, Dr. Samir R. Kapadia, Dr. Robert N. Piana, Dr. Roxana Mehran, Dr. Frederic S. Resnic, Dr. Aaron D. Kugelmass,

Dr. Sunil V. Rao, Dr. W. Douglas Weaver, and Dr. John C. Messenger.

Lastly the 16 member NCDR Management Board and 31member ACCF Board of Trustees approved these measures for submission to NQF.

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

No validity testing was necessary, other than establishing the content validity of the model, as mortality is of unquestioned importance and readily assessed.

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

We believe that the outcome is of clear importance and the construct of the risk-adjustment model has been thoroughly vetted and published in the peer-reviewed literature. Prior endorsement by NQF further supports the logic and care we used in developing this performance measure.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section* 2b4

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

The only exclusion is for patients transferred to another acute care facility, in whom their vital status cannot be readily determined.

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

There were 8,619 (<1%) patients transferred to another acute facility.

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

We do not believe that the exclusions have any impact on the validity, accuracy or interpretability of the risk-adjusted in-hospital mortality outcome measure.

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

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 40 risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

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

This is not relevant as we are proposing a risk-adjusted peri-procedural mortality outcome measure to help assess the quality and safety of PCI.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

As described in Section 2b.1.2, there was an extensive process to develop the face and contact validity of the measure. After settling on the outcome definition and candidate variables through serial conference calls with the expert panel, categorical variables were summarized as frequencies and percentages and compared with Pearson chi-squared tests. Continuous variables were summarized as medians (interquartile range) and compared using Wilcoxon rank-sum tests. Ordinal variables were tested using a chi-square test based on the rank of the group mean score.

The original model was developed by using a sample of 181,775 NCDR patients undergoing PCI from 1/04-3/06 and then validated in 2 separate samples; an additional 121,183 patients treated in the same time period and a prospective cohort of 285,440 patients treated between 3/06-3/07. Baseline patient characteristics and variables from diagnostic catheterization were considered candidate variables. Candidate variables had less than 0.5% missing data except for pre-procedure ejection fraction (29.7%). Missing values for ejection fraction were imputed by stratifying the population based on a history of congestive heart failure, prior MI, pre-procedural cardiogenic shock and the presence of STEMI to determine a median value for each patient with missing data. After the committee reviewed all variables with a statistically significant univariate association with mortality, the most clinically and statistically meaningful values were selected for potential inclusion in a logistic regression model. Backward selection with a 'stay' criterion of $p<0.05$ to develop a model predicting post-PCI mortality was then created. Variables that showed non-linear associations with the outcome were transformed using splines. All 2-way interactions were examined and significant ones were retained.

We also developed a simplified pre-procedural risk model by relying only upon pre-catheterization data that had the strongest association with mortality. This simplified model had similar discrimination and calibration and enables clinicians to estimate patients' peri-procedural mortality and share this information with patients and use the estimates to define the safest and best care for each individual patient.

The C-statistic was used to describe the discrimination of the model and replicated in clinically important subgroups of interest, including patients with and without STEMI, males and females, those aged > and ≤70 years, and patients with and without diabetes. Calibration plots were used to access goodness of fit. A p-value <0.05 was considered statistically significant. All statistical tests were two-sided. All statistical analyses were performed using SAS software (version 9.2, SAS Institute, Cary, NC).

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☐ **Internal data analysis**

☒ **Other (please describe)**

Social risk factors were not used in the risk modeling

2b3.4a. What were the statistical results of the analyses used to select risk factors?

As described above, bivariate analyses were done to identify candidate variables that differed significantly between those with and without a clinically important mortality event. Multivariable, logistic regression analyses were then performed to retain those clinically meaningful variables with a statistically significant association with mortality ($p<0.05$ for each). Table 3 in Section 1.6 demonstrates the difference between those who did and did not die after their procedure, based upon 2012 data.

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

N/A

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

The process for developing the model is described in section 2b3.3 above. Discrimination was assessed with the c-statistic and calibration was assessed with both the Hosmer-Lemeshow test and the slope of the predicted vs. observed risk. Given that the prior, approved submission included the results for the separate derivation and validation cohorts reported in Peterson et al (*J Am Coll Cardiol* 2010; 55: 1923-32), we report only the 2012 data here.

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

If stratified, skip to 2b3.9

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

The c-statistic is 0.93, which means that the probability that predicting the outcome is substantially better than chance. This method is used to compare the goodness of fit of logistic regression models. The range is between 0.5 to 1.0. A value of 0.5 indicates that the model is no better than chance at making a prediction of membership in a group and a value of 1.0 indicates that the model perfectly identifies those within a group and those not. Models are typically considered reasonable when the C-statistic is higher than 0.7. (Hosmer & Lemeshow, 2000).

The c-statistics for the original derivation and validation cohorts, as well as clinically important subgroups are provided under Table 6.

Table 6. C-Statistic Results

	Sample, n	Full Model C-Stats
Development	181,775	0.926
1st validation	121,183	0.925
2nd validation	285,440	0.924
Subgroups (in 2nd validation)		
STEMI	39,889	0.902

	Sample, n	Full Model C-Stats
No STEMI	245,551	0.892
Women	95,106	0.911
Men	190,334	0.930
Age >70 yrs	92,381	0.901
Age ≤70 yrs	193,059	0.927
Diabetes	92,974	0.924
No diabetes	192,466	0.923

Cath = catheterization; NCDR = National Cardiovascular Data Registry; STEMI = ST-segment elevation myocardial infarction.

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

The intercept for the model was -0.00063, which was not statistically significantly different than 0 (p=0.97). The slope of the calibration line was 0.9906, which also was not significantly different than 1.0 (p=0.097). A graphical representation of observed and predicted mortality rates across deciles of risk is shown under Figure 1.

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

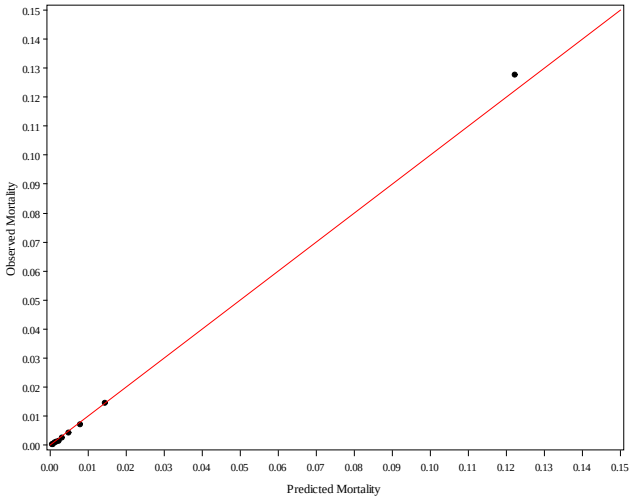


Figure 1. Calibration Curve Plot

2b3.9. Results of Risk Stratification Analysis:

The risk stratification was able to adequately segregate deciles of risk from <1% to >12% at the patient level. At the hospital level, we observed a broad range of unadjusted risk, which was substantially tightened after adjusting for patient characteristics. The unadjusted distribution of mortality is shown under Figure 2.

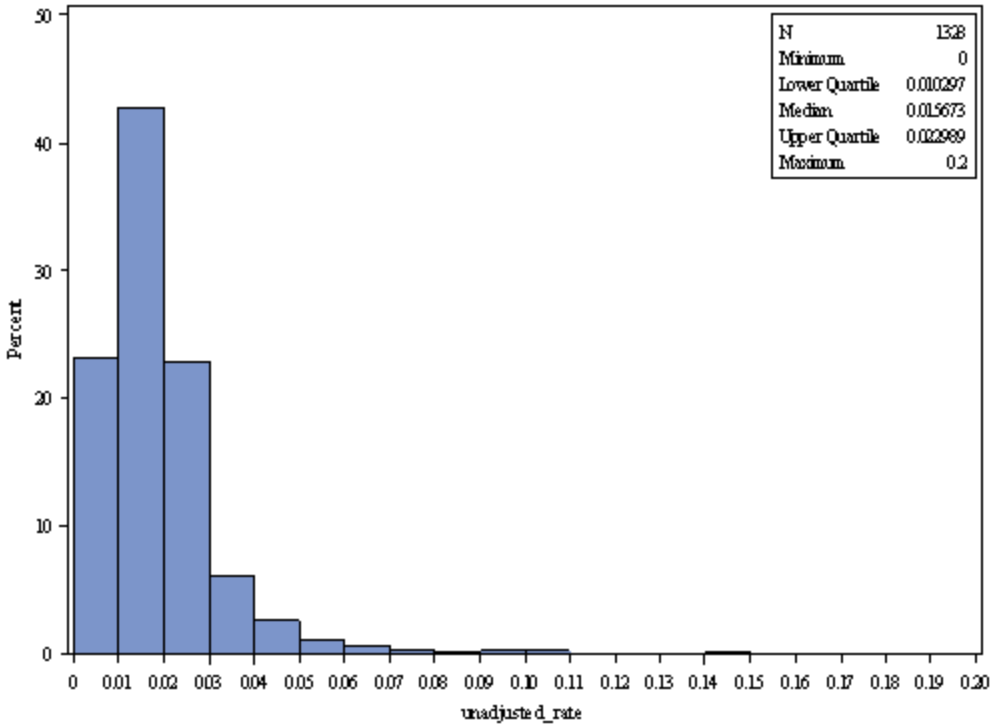


Figure 2. Unadjusted Distribution of Mortality

The mortality rates adjusted for patient characteristics is shown under Figure 3.

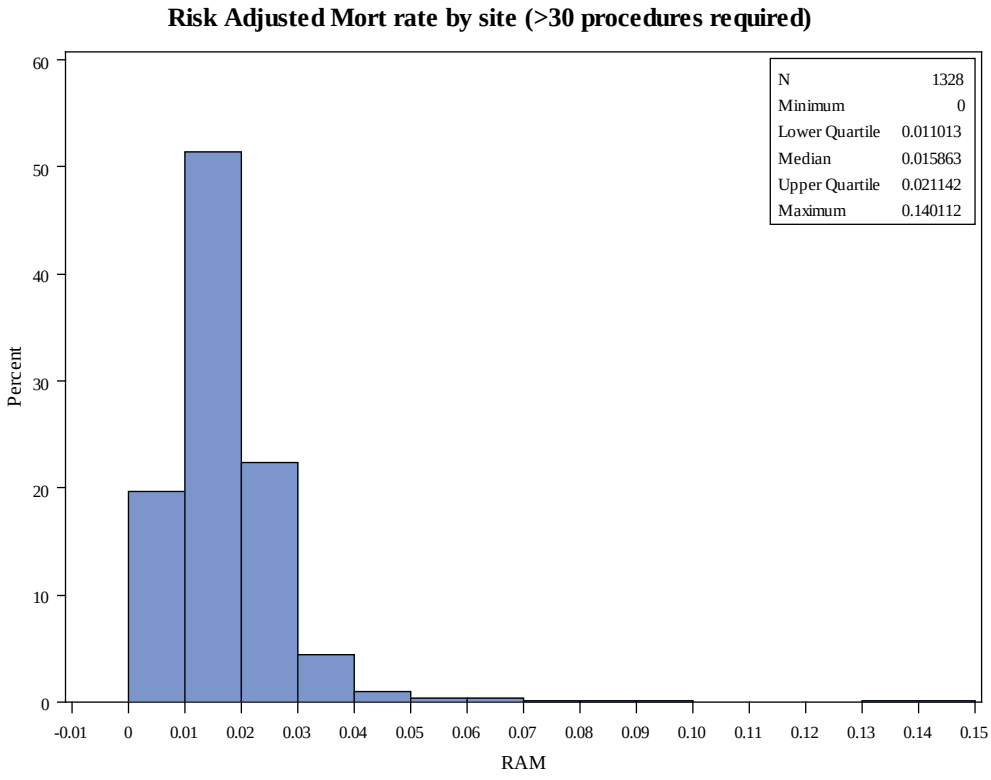


Figure 3. Adjusted Distribution of Mortality

After adjusting for patient characteristics, we observed a significantly tighter and more normal distribution of mortality events.

The distribution of sites' observed/expected ratios are shown under Figure 4.

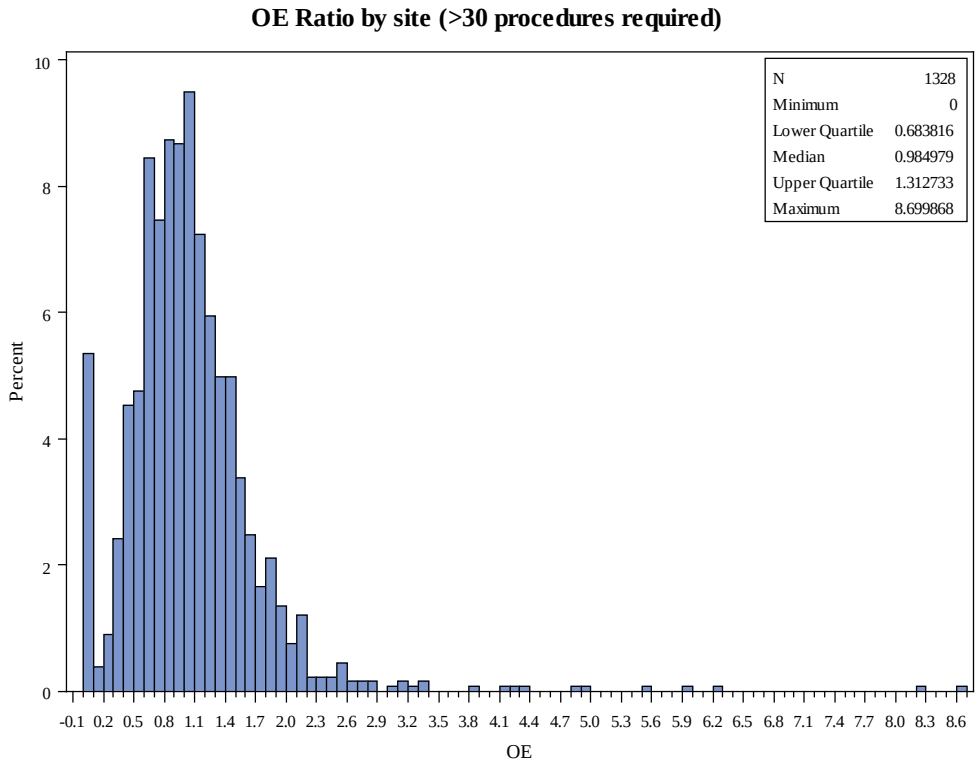


Figure 4. Site's Observed and Expected Ratio

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

We believe that the our mortality model performs exceedingly well in adjusting for patient characteristics present prior to the conduct of PCI and is able to discriminate well across a wide variety of important clinical subsets of patients. Moreover, there is substantial hospital variation before and after risk-adjusting patient characteristics. The distribution of hospitals' O/E ratios show that there are some sites with excellent performance and others with mortality rates that are more than 2-fold greater than expected. These would be sites where substantial opportunities to improve patient safety likely exist.

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

Since these data provide further strong evidence of the validity and value of the previously-endorsed measure using 2012 data, we did not do any additional testing.

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

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

As noted in the figures above, we found significant variability in mortality across hospitals, even after adjusting for pre-procedural patient characteristics. Those in the upper quartile of performance had an observed/expected ratio that was 31% greater than predicted, with some sites having a greater than 5-fold excess mortality over that predicted.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

A meaningful difference is one that indicates the potential for improvement in comparison to others. There are no absolute levels of mortality that are significant as compared with others. The average, adjusted mortality rate was 1.6% and the upper quartile ranges from 2.1 to 14%. Given an average PCI volume of 410 cases/hospital, this suggests between 2 and 50 extra deaths might be avoided per year among hospitals in the upper quartile as compared with the average hospital. Clinically, this is a large number of events, and the few significant outliers would have a very strong incentive to improve the safety of their PCI procedures.

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

We believe that the use of this model to identify outliers and the ability to pre-procedurally risk stratify patients and tailor therapy to risk holds great promise for improving the quality and safety of PCI.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

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

N/A

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

N/A

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

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

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

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

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

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

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

ACC is in the process of developing a common data dictionary mapped to coded terminology standards with the intent of improving interoperability with EHRs and potentially creation of emeasures.

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

Attachment:

3c. Data Collection Strategy

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

addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

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

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

Availability:

Participating hospitals report patient demographics, medical history, risk factors, hospital presentation, initial cardiac status, procedural details, medications, laboratory values and in-hospital outcomes as the key activity of participating in the NCDR CathPCI registry. The majority of the 17 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 data 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 by 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 NCDR CathPCI Registry. The registry is designed to include 100 percent of consecutive adult patients who undergo PCI 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 undergo a diagnostic cardiac catheterization and/or PCI. Eligible diagnostic catheterizations are characterized by the passage of a catheter into the aortic root for pressure measurements and/or angiography, and can include Left Ventricle (LV) pressure measurements, LV angiography, coronary angiography, and coronary artery bypass angiography. Eligible PCI procedures include those that involve passage or attempted passage of a coronary device across one or more coronary lesions for purposes of increasing the intraluminal diameter of the vessel and/or restoring or improving circulation. Patients are selected for inclusion by reviewing existing medical records and no direct interaction with the patient is required outside of the normal course of care. There is 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 analyzed in aggregate form without patient identifiers. The CathPCI Registry dataset, comprised of approximately 250 data elements and 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, enabling inclusion of direct identifiers in the registry to be 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. All analyses performed by contracted data analytic centers are devoid of direct patient identifiers.

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 currently include a patient survey. Physician and/or institutional confidentiality is maintained by de-identified dashboard reports. There is no added procedural risk to

patients through involvement in the CathPCI Registry. No testing, time, risk, or procedures beyond those required for routine care are 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, such as segrating all patient identifiable data from the analytic datasets provided to contracted data analytic centers.

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 are 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.

Overall there is no added procedural risk to patients through their hospital's involvement in the CathPCI Registry. No testing, time, risk, or procedures beyond those required for routine care will be imposed.

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 2017 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 CathPCI 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.

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting	Payment Program Quality Hospital Insight program for Anthem 1. https://www.anthem.com/wps/portal/ahpprovider?content_path=provider/nh/f2/s4/t0/pw_003533.htm&label=Quality Blue Distinction Centers for Cardiac Care 2. http://www.bcbs.com/healthcare-partners/blue-distinction-forproviders/cardiacprogramcriteria.pdf Quality Improvement (external benchmarking to organizations) National Cardiovascular Data Registry https://www.ncdr.com/webncdr/cathpci/home/datacollection

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

Name of program and sponsor
Quality Improvement with Benchmarking
Name of program and sponsor: Blue Distinction Centers for Cardiac Care; Sponsor: Blue Cross Blue Shield Association
Purpose:
The Blue Distinction Centers for Cardiac Care is a national designation program that recognizes hospitals that demonstrate expertise in delivering quality specialty care, safely and effectively. To earn the Blue Distinction Centers+ designation, hospitals must meet the same quality criteria as Blue Distinction Centers, and go an extra step to demonstrate that they do so cost efficiently. Quality is key: only those facilities that first meet Blue Distinction's nationally established, objective quality measures will be considered for designation as a Blue Distinction Center+. Blue Distinction Centers' goal is to help consumers find both quality and value for their specialty care needs, on a consistent basis, while encouraging healthcare professionals to improve the overall quality and delivery of care nationwide. [Retrieved from http://www.bcbs.com/healthcare-partners/blue-distinction-for-providers/cardiacprogramcriteria.pdf on 11/25/13]
Geographic area and number and percentage of accountable entities and patients included
Geographic Area: National program.
Number: Directory of Providers available at http://www.bcbs.com/why-bcbs/blue-distinction/blue-distinction-cardiac/bluedistinctioncardiac.pdf
% of accountable entities: Total of 414 hospitals
Alabama 10
Arizona 4
Arkansas 3
California 46
Colorado 6
Connecticut 5
Delaware 3

Florida 29
Georgia 4
Hawaii 1
Idaho 3
Illinois 29
Indiana 12
Iowa 8
Kansas 5
Kentucky 5
Louisiana 5
Maine 1
Massachusetts 8
Michigan 23
Minnesota 12
Missouri 12
Nebraska 5
New Hampshire 2
New Jersey 3
New York 12
Nevada 2
North Carolina 10
North Dakota 4
Ohio 26
Oklahoma 4
Patients included: information not available.
The measure is also used in the Quality Insight Hospital Program with Anthem, which overlaps with what is included above for Blue Distinction program.

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

See below

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

Update to credible plan (11/8/17):

We moved forward with implementing the in-hospital mortality measures, however, held off on public reporting since we are also in the process of updating the CathPCI registry to version 5. The new registry version includes elements to assess out-of-hospital cardiac arrest, which has been identified in the literature as a risk factor that should be considered in mortality modeling1,2. Additionally, when preparing the public reporting metric for in-hospital mortality (#0133) and 30-day mortality (#0536), we found that the measures were not harmonized in structure (i.e. the 30-day measure is a hierarchical model whereas the in-hospital measure is not). As such, these measures could not be rolled up together to create an appropriate composite view of mortality. We plan to modify the in-hospital mortality model to a hierarchical structure when we expand to take advantage of the additional elements in version 5 of CathPCI registry, particularly cardiac arrest, rather than sequencing a number of major revisions in a relatively short time period for hospitals. In order to avoid unintended negative consequences, ACC has made the decision to put a hold on public reporting until the cardiac arrest elements can be considered for modeling and the inpatient and 30-day PCI mortality models can be structurally harmonized.

Citation:

[1] Peberdy, M.A., Donnino, M.W., Callaway, C.W., et al. Impact of Percutaneous Coronary Intervention Performance Reporting on Cardiac Resuscitation Centers: A Scientific Statement From the American Heart Association. Circulation. 2013;128:762-773; originally published online July 15, 2013; doi: 10.1161/CIR.0b013e3182a15cd2

[2] Camuglia, A.C., Randhawa, V.K., Lavi, S., et al. Cardiac catheterization is associated with superior outcomes for survivors of out of hospital cardiac arrest: Review and meta-analysis. Elsevier: Resuscitation 85 (2014) 1533–1540 . www.elsevier.com/locate/resuscitation

NCDR Public Reporting Background:

ACC's National Cardiovascular Data Registry (NCDR) Voluntary Hospital Public Reporting Program: The ACC currently runs a program to give hospitals the opportunity to voluntarily publicly report their measure results based on data from the National Cardiovascular Data Registry (NCDR). Hospitals that choose to participate have their results displayed on ACC's CardioSmart. Currently Hospitals can report on five measures from the CathPCI Registry and five measures from the ICD Registry. Of these publicly reporting measures, five are NQF-endorsed:

- NQF # 1522: Use of a medicine in the ACEi or ARB class to improve heart function after ICD implant in patients with less than normal heart function.
- NQF # 1528: Use of a beta-blocker medication after ICD implant in patients with a previous heart attack.
- NQF #1529: Use of a beta-blocker medication after ICD implant in patients with less than normal heart function.
- NQF #0965: Use of all recommended medications (ACEI or ARB and beta-blocker) to improve heart function and blood pressure after ICD implant.
- NQF # 0964: Therapy with aspirin, P2Y12 inhibitor, and statin at discharge following PCI in eligible patients (composite measure)

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

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

Performance results are distributed to all CathPCI registry participants as part of quarterly benchmark reports, which provide a detailed analysis of an individual institution´s performance in comparison with the entire registry population from participating hospitals across the nation. Reports include an executive summary dashboard, at-a-glance assessments, and patient level drill-downs. Registry participants also have access to an outcome report companion guide which provides common definitions and detailed metric specifications to assist with interpretation of performance rates.

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

Results are provided as part of quarterly performance report which includes a rolling 4 quarters of data. Participating hospitals in the CathPCI registry report the following: Patient demographics for cardiac catheterization and PCI procedures, provider and facility characteristics, history/factors, cardiac status, treat lesions; intracoronary device utilization and adverse event rates; appropriate use criteria for coronary revascularization; compliance with ACC/AHA clinical guideline recommendations.

The majority of the required data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions can manually report using a free web-based tool or automate the reporting by using certified software developed by third-party vendors. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

There are a number of methods used to educate and provide general support to registry participants. These include the following:

- Registry Site Manager Calls are available for all NCDR participants. RSM calls are provided as a source of communication between NCDR and participants to provide a live chat Q and A session on a continuous basis.
- New User Calls are available for NCDR participants, and are intended for assisting new users with their questions.
- NCDR Annual Conference

The NCDR Annual Conference is a well-attended and energetic two-day program at which participants from across the country come together to hear about new NCDR and registry-specific updates. During informative general sessions, attendees can learn about topics such as transcatheter therapies, the NCDR dashboard, risk models, data quality and validation, and value-based purchasing. Attendees also receive registry updates and participate in advanced case studies covering such topics as Appropriate Use Criteria and outcomes report interpretation.

- Release notes (for outcomes reports)
- Clinical Support

The NCDR Product Support and Clinical Quality Consultant Teams are available to assist participating sites with questions Monday through Friday, 9:00 a.m. - 5:00 p.m. ET.

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

Describe how feedback was obtained.

Feedback is typically obtained through monthly registry site manager monthly calls, ad hoc phone calls tracked with salesforce software, and during registry –specific break-out sessions at the NCDR's annual meeting. Registry Steering Committee members may also provide feedback during regularly scheduled calls.

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

Registry participants have communicated to ACC that this measure is easy to understand and interpret. ACC does not receive a lot of questions about the measure and participants seem to be apply the coding instructions correctly to the data elements that impact this measure.

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

No other feedback was received from other users.

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

N/A: the measure was not modified since last endorsement.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

This measure does not readily lend itself to improvement across the entire population of hospitals over time because evolutions in technology (e.g. circulatory support) enables sicker patients to be treated. However, the measure does have the opportunity to identify hospitals with higher mortality rates than expected. This both enables hospitals to recognize this problem and develop processes to improve their performance, while also enabling external agencies

(state government and payers) to take action to either regulate the institutions (e.g. state governments) or preferentially direct their patients to hospitals with better outcomes (payers).

4b2. Unintended Consequences

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

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

Although our previous CathPCI Registry mortality models had many assets, they had been criticized for failing to accurately define risk among "extreme risk" patients, such as those with cardiogenic shock and those who have suffered cardiac arrest prior to PCI. This led to concerns about whether decision makers will adopt risk-averse patterns of patient care. In response to these concerns, and to further define risk at the highest end of the spectrum, a series of new variables were included in the 2009 updated Version 4 CathPCI Registry data clarification form (DCF v4). These variables have recently been incorporated into the CathPCI Registry risk adjustment model that is currently used for site-level outcome reporting. Model performance was assessed by discrimination and calibration metrics in a separate split sample. In-hospital mortality was 1.4%, ranging from 0.2% among elective cases (45.1% of total cases) to 65.9% among patients with shock and recent cardiac arrest (0.2% of total cases). With the inclusion of indicators for high-risk PCI, the updatedCathPCI Registry DCF v4 mortality models perform well in both low- and high-risk PCI patient populations.[1] There have been no significant concerns raised about the current adequacy of risk adjustment and the inclusion of the additional data elements to better account for patient severity seem to have satisfactorily met the concerns of the interventional community. Accordingly, all elements of the current risk model are being retained in the planned release of an upcoming data collection form, CathPCI Version 5. In addition, cardiac arrest data elements identified in the literature as risk factors will be included in the new version of the registry. We plan to update the risk model to accommodate these elements accordingly.

[1] Brennan J, Curtis JP, Dai D, et al. Enhanced Mortality Risk Prediction With a Focus on High-Risk Percutaneous Coronary Intervention: Results From 1,208,137 Procedures in the NCDR (National Cardiovascular Data Registry). J Am Coll Cardiol Intv. 2013;6(8):790-799. doi:10.1016/j.jcin.2013.03.020. Retrieved from http://interventions.onlinejacc.org/article.aspx?articleid=1730158

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

We are unaware of any unanticipated benefits or harms from the implementation of this measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0119 : Risk-Adjusted Operative Mortality for CABG

0230 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older

0535 : 30-day all-cause risk-standardized mortality rate following percutaneous coronary intervention (PCI) for patients without ST segment elevation myocardial infarction (STEMI) and without cardiogenic shock

0536 : 30-day all-cause risk-standardized mortality rate following Percutaneous Coronary Intervention (PCI) for patients with ST segment elevation myocardial infarction (STEMI) or cardiogenic shock

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

No

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

Measure 119 offers a risk-adjusted measure for mortality, as does our Risk-Adjusted Mortality measure. The patient population is similar in that both these measures evaluate the mortality for patients requiring coronary artery revascularization. The measure stewarded by STS provides a risk adjusted outcome evaluated at 30 days post their CABG surgery. While the NCDR measure evaluates mortality at discharge from the index admission for the PCI. The method of revascularization differs between the two measures, rendering the overlap insubstantial.

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.)

As noted in the previous section, these measures focus on different populations or have different durations of follow-up (30 days vs. in-hospital). We believe that because PCI is the most common cardiac procedure for coronary artery disease, is associated with substantial costs and is variable across hospitals, that there is great importance in having a measure specifically devoted to the outcomes of this procedure.

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: Enhanced_mortality_risk_prediction_with_a_focus_on_high_-risk_PCi-636426313189114692.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American College of Cardiology

Co.2 Point of Contact: Esteban, Perla, eperla@acc.org, 202-375-6499-

Co.3 Measure Developer if different from Measure Steward: American College of Cardiology

Co.4 Point of Contact: Kim, Lavin, klavin@acc.org, 202-375-6448-

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 interventional cardiology. 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, Issam Moussa, and David Malenka.

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: 2001

Ad.3 Month and Year of most recent revision: 04, 2012

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? 04, 2018

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 (cardiosource.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.

Ad.8 Additional Information/Comments: ACC appreciates the opportunity to submit measures for this NQF endorsement maintenance project.

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0536

Corresponding Measures:

De.2. Measure Title: 30-day all-cause risk-standardized mortality rate following Percutaneous Coronary Intervention (PCI) for patients with ST segment elevation myocardial infarction (STEMI) or cardiogenic shock

Co.1.1. Measure Steward: American College of Cardiology

De.3. Brief Description of Measure: This measure estimates hospital risk-standardized 30-day all-cause mortality rate following percutaneous coronary intervention (PCI) among patients who are 18 years of age or older with STEMI or cardiogenic shock at the time of procedure. The measure uses clinical data available in the National Cardiovascular Data Registry (NCDR) CathPCI Registry for risk adjustment. For the purpose of development and testing, the measure cohort was derived in a Medicare fee-for-service (FFS) population of patients 65 years of age or older with a PCI. For the purpose of maintenance, the measure used a cohort of patients whose vital status was determined from the National Death Index (which reflects an all-payor sample as opposed to only the Medicare population). This is consistent with the measure's intent to be applicable to the full population of PCI patients.

1b.1. Developer Rationale: This measure will describe hospital-level mortality rates following PCI in patients with STEMI or cardiogenic shock, with the overriding goal to reduce 30-day mortality rates to best-in-class. The expectation is that providing this information to hospitals, coupled with public reporting of hospitals' results, will drive internal hospital quality improvement efforts to focus efforts on reducing PCI mortality. Of note, the measure includes not only in-hospital deaths, but also deaths occurring after hospital discharge. This perspective may motivate hospitals to look for opportunities not only within the organization, but also to better coordinate the transition of care from the inpatient to the outpatient arena.

S.4. Numerator Statement: The outcome for this measure is all-cause death within 30 days following a PCI procedure in patients with STEMI or cardiogenic shock at the time of the procedure.

S.6. Denominator Statement: The target population for this measure includes inpatient and outpatient hospital stays with a PCI procedure for patients at least 18 years of age, with STEMI or cardiogenic shock at the time of procedure, including outpatient and observation stay patients who have undergone PCI but have not been admitted. It is unlikely that patients in this cohort would not be admitted to the hospital, but we keep this criterion to be consistent with the complementary non-STEMI, non-cardiogenic shock PCI cohort.

S.8. Denominator Exclusions: Hospital stays are excluded from the cohort if they meet any of the following criteria:

(1) PCIs that follow a prior PCI in the same admission (either at the same hospital or a PCI performed at another hospital prior to transfer).

This exclusion is applied in order to avoid assigning the death to two separate admissions.

(2) For patients with inconsistent or unknown vital status or other unreliable data (e.g. date of death precedes date of PCI);

(3) Subsequent PCIs within 30-days. The 30-day outcome period for patients with more than one PCI may overlap. In order to avoid attributing the same death to more than one PCI (i.e. double counting a single patient death), additional PCI procedures within 30 days of the death are not counted as new index procedures.

(4) PCIs for patients with more than 10 days between date of admission and date of PCI. Patients who have a PCI after having been in the hospital for a prolonged period of time are rare and represent a distinct population that likely has risk factors related to the hospitalization that are not well quantified in the registry.

De.1. Measure Type: Outcome

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

S.20. Level of Analysis: Facility, Other

IF Endorsement Maintenance – Original Endorsement Date: Aug 05, 2009 **Most Recent Endorsement Date:** Sep 08, 2014

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? This measure is most similar to and paired with 30-day all-cause risk-standardized mortality rate following percutaneous coronary intervention (PCI) for patients without ST segment elevation myocardial infarction (STEMI) and without cardiogenic shock. Its complementary value stems from the target population of STEMI and/or shock patients.

Staff Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Summary of prior review in 2014

- In their rationale, the developer references literature supporting an association with improved survival and the use of preprocedural clopidogrel and glycoprotein 2b/3a inhibitors; the volume of iodinated contrast; and participation in continuous quality improvement programs.
- The Committee agreed that the importance of the outcome is self-evident.

Changes to evidence from last review

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

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

Updates: The developer stated that there are no updates to the evidence.

The developer provided performance data from 1,276 hospitals and 94,907 admissions from 2011-2014 demonstrating a variation in risk-standardized mortality rates with a range from 4.7% to 15.7%.

Empirical data demonstrating a relationship between the outcome to at least one healthcare process is now required. NQF guidance states that a wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.

Question for the Committee:

- *Does the stated rationale link lower mortality rates after PCI to at least one healthcare action?*
- *Is the performance data sufficient, in size and variance, to demonstrate that some hospitals are engaging in quality improvement activities to decrease mortality after PCI better than others?*

Guidance from the Evidence Algorithm

Health outcome measure (Box 1) -> relationship between the measured health outcome and at least one healthcare action is demonstrated -> Pass

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided the combined risk-standardized mortality rates for all payers and all ages (>18 years) from ~~1,276~~ 1,356 hospitals and ~~94,907~~ 245,877 admissions using NCDR CathPCI data linked with National Death Index (NDI) from 2011-2014. The developer noted that October 2012 and November 2012 data was excluded due to missing data.
- The developer also provided the following RSMR data:
  - Mean: 8.3%
  - Standard Deviation: 1.6%
  - Range: 4.7% to 15.7%
  - IQR: 7.3% to 9.3%
- The range of performance on this data set is:

Percentile of RSMR	Mean RSMR
100% Max	0.1566
99%	0.1252
95%	0.1127
90%	0.1046
75% Q3	0.0932

Percentile of RSMR	Mean RSMR
50% Median	0.0812
25% Q1	0.0725
10%	0.0646
5%	0.0604
1%	0.0538
0% Min	0.0469

- The range by year is:

Percentile of RSMR	2011-12	2012-13	2013-14
100% Max	0.1487	0.1504	0.2013
99%	0.1346	0.1307	0.1434
95%	0.1127	0.1147	0.1204
90%	0.1025	0.1067	0.1101
75% Q3	0.0885	0.0937	0.0958
50% Median	0.0780	0.0825	0.0850
25% Q1	0.0681	0.0736	0.0742
10%	0.0603	0.0647	0.0667
5%	0.0565	0.0593	0.0628
1%	0.0505	0.0520	0.0543
0% Min	0.0459	0.0438	0.0453

Disparities:

- The developer provided the following hospital-level RSMR disparities data by race and hospital safety net status:

  Distribution of 30-day RSMR for STEMI/Shock Stratified by Quartile of Non-White Patients from 2011-2014

Description	RSMRs by Hospital Quartile of Non-White Patients			
	Q1	Q2	Q3	Q4
N	341	337	339	339
Mean	0.0831	0.0845	0.0856	0.0848
SD	0.0134	0.0155	0.0153	0.0141
100% Max	0.1459	0.1566	0.1342	0.1284
50% Median	0.0820	0.0828	0.0829	0.0828
0% Min	0.0509	0.0469	0.0506	0.0539

	Safety Net Hospitals	Non-Safety Net Hospitals
Median RSMR	8.4%	8.2%
Interquartile Range	7.4% to 9.3%	7.6% to 9.5%

Questions for the Committee:

- *Does the measure demonstrate a quality problem related to mortality in patients undergoing PCI?*
- *Is a national performance measure still warranted?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

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

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

1a. Evidence:

- This measure was first endorsed in 2009 and reviewed again in 2014. There has been no additional evidence reported.
- No changes to evidence for this maintenance measure
- Evidence was considered adequate in 2014
- Published studies through the ACC have demonstrated the 30-day mortality rates correlate with the quality of care received in the hospital for management of patients with STEMI or cardiogenic shock. This adds to the relevance of the measure.
- Pass rating for evidence with evidence unchanged since last evaluation.
- With this maintenance measure the evidence is directly related to the outcome and and their has not been an update to the evidence.
- pass

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

- Data were submitted by the developer. The data supports continued use of the measure
- There was a modest performance gap ranging from 4.7%-15.7%.
- There is a two-fold difference in mortality between the 5th and 95th %ile. Opportunity for improvement is moderate.
- Studies published in Circulation still demonstrate a gap in outcomes suggesting that this is still an important measure. See

  Development of 2 registry-based risk models suitable for characterizing hospital performance on 30-day all-cause mortality rates among patients undergoing percutaneous coronary intervention.

  Curtis JP, Geary LL, Wang Y, Chen J, Drye EE, Grosso LM, Spertus JA, Rumsfeld JS, Weintraub WS, Masoudi FA, Brindis RG, Krumholz HM. Circ Cardiovasc Qual Outcomes. 2012 Sep 1;5(5):628-37.
- Moderate performance gap and opportunity for improvement. Lack of data for 2014-2017. RSMR seems to be getting worse, therefore is quality measure having any effect? Disparities addressed, but more data would be beneficial.
- The performance gap is significant and represents an opportunity for significant performance/outcome improvement given the degree of variability.
- medium level

Criteria 2: Scientific Acceptability of Measure Properties

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: Staff

Evaluation of Reliability and Validity (and composite construction, if applicable): Link to Scientific Acceptability

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Scientific Acceptability

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.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 0536

Measure Title: 30-day all-cause risk-standardized mortality rate following Percutaneous Coronary Intervention (PCI) for patients with ST segment elevation myocardial infarction (STEMI) or cardiogenic shock

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
   NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.

   TIPS: Consider the following: 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?

   ☒Yes (go to Question #2)

   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.
2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

   TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

   ☒Yes (go to Question #4)

   ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)
3. Was **empirical VALIDITY testing** of patient-level data conducted?

   ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

   ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)
4. Was reliability testing conducted with computed performance measure scores for each measured entity?

   TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

   ☒Yes (go to Question #5)

   ☐No (go to Question #8)

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

   TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

   ☒Yes (go to Question #6) A test-retest approach was performed with an ICC of 0.122.

   ☐No (please explain below then go to Question #8)

6. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

   TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?

   Do the results demonstrate sufficient reliability so that differences in performance can be identified?

   ☐High (go to Question #8)

   ☒Moderate (go to Question #8)

   ☐Low (please explain below then go to Question #7)

7. Was other reliability testing reported?

   ☐Yes (go to Question #8)

   ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

8. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

   TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

   ☐Yes (go to Question #9)

   ☐No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)

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

   TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

   Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

   ☐Yes (go to Question #10)

   ☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

10. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

    ☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

    ☐Insufficient (go to Question #11)

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

1. Were all potential threats to validity that are relevant to the measure empirically assessed?

   TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

   ☒Yes (go to Question #2)

   ☐No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

2. Analysis of potential threats to validity: Any concerns with measure exclusions?

   TIPS: Consider the following: 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)? If patient preference (e.g., informed decision-making) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

   ☐Yes (please explain below then go to Question #3)

   ☒No (go to Question #3)

   ☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

3. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

   ☐Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

   a. Is a conceptual rationale for social risk factors included? ☒Yes ☐No

   b. Are social risk factors included in risk model? ☐Yes ☒No

   The developer noted that given the clinical data available, social risk factors (which are not readily available) would not likely contribute much improvement to this particular risk model.

   c. Any concerns regarding the risk-adjustment approach?

   *TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? 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, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

   ☐Yes (please explain below then go to Question #4)

☒No (go to Question #4)

4. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

   ☐Yes (please explain below then go to Question #5)

   ☒No (go to Question #5)

5. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

   ☐Yes (please explain below then go to Question #6)

   ☐No (go to Question #6)

   ☒Not applicable (go to Question #6)

6. Analysis of potential threats to validity: Any concerns regarding missing data?

   ☐Yes (please explain below then go to Question #7)

   ☒No (go to Question #7)

ASSESSMENT OF MEASURE TESTING

7. Was empirical validity testing conducted using the measure as specified and appropriate statistical test?

   Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

   ☒Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

   ☐No (please explain below then go to Question #8)

8. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

   TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

   ☐Yes (go to Question #9)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

9. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

   ☐Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

10. Was validity testing conducted with computed performance measure scores for each measured entity?

    TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

    ☒Yes (go to Question #11)

☐No (please explain below and go to Question #13)

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

    TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

    ☒Yes (go to Question #12)

    ☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

12. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #14)

    ☒Moderate (go to Question #14)

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

13. Was other validity testing reported?

    ☒Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

14. Was validity testing conducted with patient-level data elements?

    TIPS: Prior validity studies of the same data elements may be submitted

    ☒Yes (go to Question #15)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

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

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #16)

    ☒No (please explain below and rate Question #16 as INSUFFICIENT)

16. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

    ☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

    ☒Insufficient (go to Question #17)

    The developer provided an the overall percent agreement score; NQF guidance states that all critical data elements must be assessed separately.

17. OVERALL VALIDITY RATING

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

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

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

2a1. Specifications:

- The specs are reasonable
- The data elements have been used over the past 8 years and have demonstrated reliability and reproducibility
- Measure specifications are defined and consistent with the evidence.
- The only concern here is the inclusion of STEMI and cardiogenic shock. although related they are not necessarily synonymous. This means that although a small percentage of patients could have a mutually exclusive diagnosis- this would represent 2 different populations being measured.
- medium level of reliability

2a2. Reliability testing:

- No concerns. Rated as moderate reliability.
- The test-retest within the same hospital had a somewhat low agreement.
- Reliability is moderate.
- No
- Reliability testing was done through a test-retest approach with an ICC of 0.122, which is defined as slight, attributed to the low number of cases in the cohort per the developer. Low to moderate reliability.
- only as above
- medium level

2b2. Validity testing & 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data):

- Rated at moderate validity.
- data element validity was generally high
- Validity is moderate without significant threats.
- The analyses provided in the papers do demonstrate differences which appear to be significant and meaningful There has been improvement over the years so that the variations are less clear than initially
- Empirical validity testing done through data element validity testing with agreement results at 92%, indicating moderate to high correlation. Overall moderate to high validity.
- since all cause mortality is used as the endpoint and direct attribution to STEMI or complications of PCI do not have to be defined, the validity of the metric should be high.
- have concerns with the testing results. do not feel this passes due to lack of information

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

- risk adjustment model had a c-statistic of 0.825
- The measure is risk adjusted. The method passed in 2014
- Apparently CMS data is not accessible to ACC for data calculations. This may introduce errors if certainly not bias for the lack of the most vulnerable part of the population
- Exclusions noted and clear with no apparent threats. Risk adjustment done with statistical risk model, and results demonstrate good model discrimination. Data from risk model is 2010-2011-might need updating. Social risk factors included, but not included in risk model due to not being readily available and unlikely to contribute to improvement.
- not addressed

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent
3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer states:

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
- ACC is in the process of developing a common data dictionary mapped to coded terminology standards with the intent of improving interoperability with EHRs and potentially creation of eMeasures.
- Implementation of this measure requires matching of patient data to external data source to determine the outcome endpoint (Death 30 days after PCI). This has resulted in several implementation challenges
    - Data Availability: ACC is not able to use CMS data as a source for this measure as it is not being used for research purposes (CMS ResDAC path) and they do not have other payor data after the hospital visit (Qualified Entity requirement) and do not fit either path to receive CMS Data. They have had to change implementation strategy, rework their models and match NCDR records to CDC National Death Index (NDI) data.
    - Patient Confidentiality: CDC NDI requires direct patient identifiers in order to meet the minimum criteria for matching. Roughly 15% of submitting NCDR sites (based on 2017Q2 CathPCI data) do not submit direct patient identifiers to the registry and are therefore ineligible for NDI matching and cannot participate in this measure.
    - Data Cost: CDC NDI charges for matching of data to NDI. This results in a $100,000 cost to ACC per year to report this measure for the CathPCI registry.
    - Data Timeliness: CDC NDI is released on a yearly basis, roughly one year after the calendar year of death along with processing time of the matching process and report generation and the most contemporary data available is over

18 months old. In addition, calculating 30 Day mortality for December of the calendar year requires waiting for an extra 12 months (30 months total) in order to get death certificates for January of the following year.

- There are fees and licensing requirements for use of this measure in the ACC's National Cardiovascular Data Registry (NCDR) program.

Questions for the Committee:

- o *Are the required data elements routinely generated and used during care delivery?*
- o *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- o *Considering the implementation challenges faced by the developer, do you have any concerns about the feasibility of the measure?*

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

Committee Pre-evaluation Comments: Criteria 3: Feasibility

3. Feasibility

- This measure has demonstrated that it is feasible.
- outcome determined from another data source, such as the National Death Index.
- Feasibility is moderate. The developer is moving toward making this an e-measure
- This is an ACC tool so that data from non-cardiologists may not be included. However this has been used for several years
- Data collection obtained through electronic sources via administrative claims data as well as registry data. Several data implementation challenges identified, including availability, cost, and patient confidentiality. Moderate feasibility.
- ACC is "in process" of standardizing terminology in their dictionary which could cause logistic problems until finalized. Also, there is a cost to use the ACC registry which may not be budgeted by all facilities.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details: None

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

measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others: The developer states that performance results are distributed to all CathPCI registry participants as part of benchmark reports, which provide a detailed analysis of an institution´s individual performance in comparison to the entire registry population from participating hospitals across the nation.

Additional Feedback: Feedback is typically obtained through monthly registry site manager monthly calls, ad hoc phone calls tracked with salesforce software, and during registry –specific break-out sessions at the NCDR's annual meeting. Registry Steering Committee members may also provide feedback during regularly scheduled calls. The developer states that since it was implemented relatively recently, there have been no major issues or other feedback received from registry participants with respect to collecting data for this particular metric.

***Questions for the Committee**:*

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

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

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

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

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

Improvement results: The developer stated they were unable to comment on or draw conclusions from risk adjusted performance trends over time because different cohorts of data (CMS vs NDI) were analyzed.

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

Unexpected findings (positive or negative) during implementation: Implementation challenges were outlined in the feasibility section as the availability, cost, and timeliness of data and concerns about patient confidentiality.

Potential harms: The developer states studies suggest that public reporting of the outcomes of cardiovascular procedures may have unintended consequences but determining the underlying causes and appropriateness of these differences is not possible at this time. This measure has not undergone public reporting to date, thus the unintended consequences are speculative.

Additional Feedback: None

***Questions for the Committee**:*

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

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

Rationale: Trend data has not been provided so it is not possible to determine if progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is occurring. The developer has experienced several implementation challenges and there is a risk of harm to the patient if this measure is publically reported.

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

4a. Use

- The measure is publicly reported but it is unlear if it is used in an accountability program.
- Used for quality improvement

- The measure is in use but not in an accountability program. The results are distributed to the users of the data base.
- As commented upon in a JACC editorial, the use of PCI may be a ""by-stander"" event and the care for the PCI may not be causally related to mortality. JACC: Cardiovascular Interventions, Volume 9, Issue 5, 14 March 2016, Pages 496-498"
- Measure is currently reported publically ACC NCDR. Additional reporting recently in CathPCI registry. Thus, pass on use.
- Although publicy reported, no accountability program is currently being used. Also, feedback relies on anecdotal and non structured input received through site visits and at meetings. There is not a FORMAL process in place to illicit feedback and to collect this data in a structured way.

4b. Usability

- There is concern expressed by the developer that there may be harm with public reporting but there is no evidence that has happened.
- Trend data has not been provided so it is not possible to determine if progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is occurring. The developer has experienced several implementation challenges and there is a risk of harm to the patient if this measure is publicly reported.
- The primary harm is if there is a misinterpretation of the data. So the results must be carefully explained to assure no unintended consequences with either over-confidence in a facility's results or with over-concern as to anticipated outcomes.
- However, because quality improvement cannot be determine from trend data, the existence of several implementation challenges, and the potential for patient harm from the unintended consequences of public reporting, usability is insufficient.
- This metric could be used to improve policies and procedures by incorporating standard best practices. However, the patient pool for this measure is inherently high risk. Therefore, patient selection or deselection could be an unintended consequence based on risk stratification. This could potentially delay care for patients.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0229 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older
- 0230 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older
- 0535 : 30-day all-cause risk-standardized mortality rate following percutaneous coronary intervention (PCI) for patients without ST segment elevation myocardial infarction (STEMI) and without cardiogenic shock

Harmonization

- The developer stated that this measure is as harmonized as possible to the related and competing measures.

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

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 10, 2018

No comments have been submitted as of this date.

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

0536_NQF_evidence_attachment_20171108.pdf

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission? Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

1a Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0536

Measure Title: 30-day all-cause risk-standardized mortality rate following Percutaneous Coronary Intervention (PCI) for patients with ST segment elevation myocardial infarction (STEMI) or cardiogenic shock

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: Click here to enter a date11/8/2017

Instructions

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

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

1a. Evidence to Support the Measure Focus

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

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.

- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

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

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

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

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

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

Outcome

☒ Outcome: 30-day mortality

☐Patient-reported outcome (PRO):

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

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

☐ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The goal of this measure is to reduce PCI 30-day mortality rates to best-in-class. 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.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

Evidence that the outcome measure has been influenced by one or more clinical interventions:

Numerous studies have demonstrated the efficacy of interventions designed to improve patient outcomes following PCI. These include pharmacologic interventions such as the use of glycoprotein 2b/3a inhibitors, direct thrombin inhibitors, and pre-procedural thienopyridines such as clopidogrel and prasugrel, as well as advances in device technology such as use of stents,, thrombectomy for acute lesions with high thrombus burden, and distal embolic protection for PCI of degenerated saphenous vein grafts. Of note, the majority of these interventions have been shown to reduce endpoints other than mortality, most commonly rates of periprocedural MI, major bleeding, and target vessel revascularization for in-stent restenosis. Although few individual interventions have been shown to reduce mortality, they may collectively exert a favorable impact on hospital PCI mortality rates when implemented in a coordinated fashion.

There is a growing body of evidence that quality improvement efforts can improve outcomes of PCI patients, including survival. Rihal and colleagues examined patient outcomes before and after initiation of a program of continuous quality improvement (CQI) and found a significantly lower in-hospital mortality following PCI despite significant increases in the risk profile of PCI patients. Similar improvements were identified in studies of CQI by Brush et al and Moscucci et al, and improvements in survival were associated with greater adherence to evidence-based practices including preprocedural clopidogrel, use of glycoprotein 2b/3a inhibitors, and volume of iodinated contrast. The observational nature of these studies precludes drawing definitive conclusions, but they strongly suggest a mechanism through which public reporting of hospital PCI outcomes could promote improvements in the care of PCI patients.

References:

Brush JE, Balakrishnan SA, Brough J, Hartman C, Hines G, Liverman DP, Parker JP, Rich J, Tindall N. (2006). "Implementation of a continuous quality improvement program for percutaneous coronary intervention and cardiac surgery at a large community hospital." Am Heart J 152 (2):379-85 16875926 (P,S,E,B).

Krumholz HM, Brindis RG, 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(3): 456-62.

Moscucci M, Kline Rogers E, Montoye C, Smith DE, Share D, O'Donnell M, Maxwell-Eward A, Meengs WL, De Franco AC, Patel K, McNamara R, McGinnity JG, Jani SM, Khanal S, Eagle KA. (2006). "Association of a Continuous Quality Improvement Initiative With Practice and Outcome Variations of Contemporary Percutaneous Coronary Interventions." Circulation. 113:814-822.

Rihal C, Kamath C, Holmes D, et al. (2006). "Economic and clinical outcomes of a physician-led continuous quality improvement intervention in the delivery of percutaneous coronary intervention." Am J Manag Care 12:445-452.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

See evidence/literature described above in 1a3.

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

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

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

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

This measure will describe hospital-level mortality rates following PCI in patients with STEMI or cardiogenic shock, with the overriding goal to reduce 30-day mortality rates to best-in-class. The expectation is that providing this information to hospitals, coupled with public reporting of hospitals' results, will drive internal hospital quality improvement efforts to focus efforts on reducing PCI mortality. Of note, the measure includes not only in-hospital deaths, but also deaths occurring after hospital discharge. This perspective may motivate hospitals to look for opportunities not only within the organization, but also to better coordinate the transition of care from the inpatient to the outpatient arena.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

The study cohort for the validation of this measure includes NCDR CathPCI data linked with National Death Index (NDI) data to ascertain the specifications for 30-day RSMRs for all payers and all ages (>18 years). Using the previously endorsed measure (there have been no changes to the specifications), we analyzed variation in 30-day RSMRs among the hospitals in this linked dataset for a three-year period, from December 2011 to December 2014. We excluded two months of observation due to missing data during our sampling frame (October 2012 and November 2012). There were ~~94,907~~ 245,877 admissions to ~~1,276~~ 1,356 hospitals in the combined three-year sample. RSMRs varied among hospitals, with a mean of 8.3%, a standard deviation of 1.6%, and a range of 4.7% to 15.7%. The interquartile range was 7.3% to 9.3%. The range of performance is as follows:

Percentile of RSMR	Mean RSMR
100% Max	0.1566
99%	0.1252
95%	0.1127
90%	0.1046
75% Q3	0.0932
50% Median	0.0812
25% Q1	0.0725
10%	0.0646
5%	0.0604
1%	0.0538
0% Min	0.0469

Below is a histogram of the distribution of 30-day RSMR for STEMI/Shock:

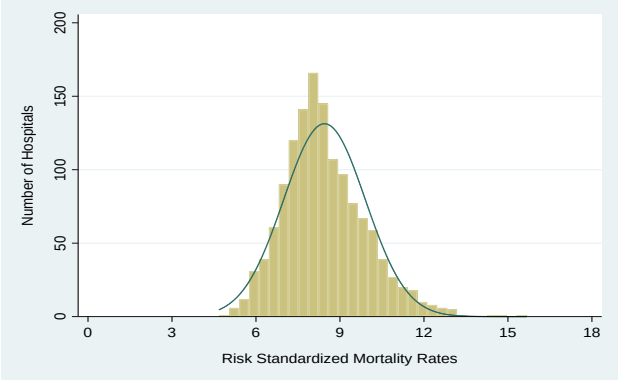
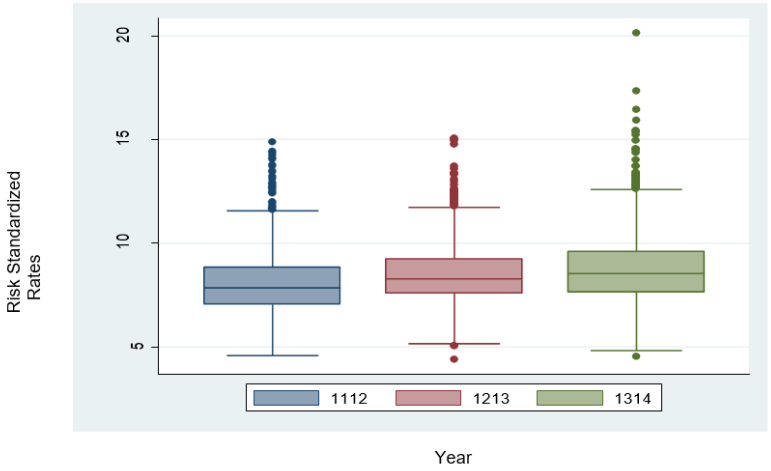


Table 2: Distribution of hospital 30-day RSMR for STEMI/Shock by year

Description and Percentile	**Mean RSMR by Year**		
	2011-2012	**2012-2013**	**2013-2014**
N	74067	76903	94907
Mean	0.0800	0.0844	0.0869
Std Deviation	0.0170	0.0165	0.0179
100% Max	0.1487	0.1504	0.2013
99%	0.1346	0.1307	0.1434
95%	0.1127	0.1147	0.1204
90%	0.1025	0.1067	0.1101
75% Q3	0.0885	0.0937	0.0958
50% Median	0.0780	0.0825	0.0850
25% Q1	0.0681	0.0736	0.0742
10%	0.0603	0.0647	0.0667
5%	0.0565	0.0593	0.0628
1%	0.0505	0.0520	0.0543
0% Min	0.0459	0.0438	0.0453

Figure 2: Boxplot of hospital 30-day RSMR for STEMI/Shock, by year



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.

Inpatient mortality is the indicator that has been most widely used to evaluate the quality of cardiac procedures and is arguably the most important adverse outcome measure. The ACC summarized the experience of the NCDR CathPCI Registry from 1998-2000 and found that in-hospital mortality occurred in 1,422 of 100,253 PCI procedures (1.4%) (Shaw, Anderson et al. 2002). Mortality was higher in patients with acute myocardial infarction (4.9%) or cardiogenic shock (27.2%). In the present era, mortality rates for PCI in large series from experienced operators varied across hospitals (Carrozza, Cutlip et al. 2008). Prior studies have demonstrated significant variability in in-hospital PCI mortality across age groups, gender, geographic regions, socioeconomic status, and by hospital volume (Mukherjee, Wainess et al. 2005). Although 12 states already report PCI outcomes, to date there has not been a unified national effort to publicly report PCI mortality.

Citations

Carrozza J, Cutlip D, Levin T. (2008). Periprocedural complications of percutaneous coronary intervention. UpToDate. B. Rose. Waltham, MA.

Mukherjee D, Wainess RM, et al. (2005). "Variation in outcomes after percutaneous coronary intervention in the United States and predictors of periprocedural mortality." Cardiology 103(3): 143-7.

Shaw RE, Anderson HV, et al. (2002). "Development of a risk adjustment mortality model using the American College of Cardiology-National Cardiovascular Data Registry (ACC-NCDR) experience: 1998-2000." J Am Coll Cardiol 39(7): 1104-12.

Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, Hailpern SM, Ho M, Howard V, Kissela B, Kittner S, Lloyd-Jones D, McDermott M, Meigs J, Moy C, Nichol G, O'Donnell C, Roger V, Sorlie P, Steinberger J, Thom T, Wilson M, Hong Y. Heart Disease and Stroke Statistics_2008 Update: A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee and for the American Heart Association Statistics Committee and Stroke Statistics Subcommittee Circulation 2008;117;e25-e146; originally published online Dec 17, 2007; DOI: 10.1161/CIRCULATIONAHA.107.187998.

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

We analyzed whether disparities in performance on this measure exist at the hospital-level by race and hospital safety set status.

To identify potential disparities by race, we examined the relationship between hospital-level RSMR and hospital proportion of non-White patients among all hospitals grouped by quartile of the proportion of non-White patients.

Analyses demonstrated that the median RSMR for hospitals with the highest quartile of non-White patients was 8.3% compared with 8.2% among hospitals with the lowest quartile of non-White patients. The distributions for the RSMRs overlapped, and many hospitals caring for the highest quartile of non-White patients performed well on this measure. In addition, in comparison to the registry mean RSMR of 8.3%, hospitals with the highest proportions of non-White patients do not have worse 30-day RSMRs in the CathPCI-NDI linked cohort.

Distribution of 30-day RSMR for STEMI/Shock Stratified by Quartile of Non-White Patients

Description	RMSRs by Hospital Quartile of Non-White Patients			
	Q1	Q2	Q3	Q4
N	341	337	339	339
Mean	0.0831	0.0845	0.0856	0.0848
Std Deviation	0.0134	0.0155	0.0153	0.0141
100% Max	0.1459	0.1566	0.1342	0.1284
99%	0.1218	0.1289	0.1253	0.1252
95%	0.1048	0.1127	0.1146	0.1108
90%	0.1001	0.1042	0.1068	0.1042
75% Q3	0.0909	0.0930	0.0952	0.0926
50% Median	0.0820	0.0828	0.0829	0.0828
25% Q1	0.0741	0.0743	0.0746	0.0754
10%	0.0672	0.0664	0.0674	0.0680
5%	0.0630	0.0615	0.0650	0.0636
1%	0.0571	0.0568	0.0551	0.0575
0% Min	0.0509	0.0469	0.0506	0.0539

Similarly, to identify potential disparities related to socioeconomic status (SES), we examined the relationship between RSMR and hospital safety net status. Safety net status was defined as government (public) hospitals or non-government hospitals with a caseload that is higher than the average of the Medicaid caseloads of hospitals within a given state plus one standard deviation of Medicaid caseload of hospitals within that state. We used the American Hospital Association data (2010) to calculate the Medicaid caseload and define hospital safety net status (Yes/No). Hospital safety net status was used as a marker of SES because safety net hospitals serve a low income and vulnerable patient population.

Analyses demonstrated that the median RSMR was 8.4% for safety net hospitals compared with 8.2% for non-safety net hospitals. The interquartile range for safety net hospitals was 7.4% to 9.3%, whereas among non-safety net hospitals it was 7.6% to 9.5%. Overall, hospitals with a high proportion of vulnerable patients, as defined by safety net status, do not have worse 30-day RSMRs in this cohort.

Consistent with NQF guidelines, this measure does not risk adjust for race or SES.

Distribution of 30-day RSMR for STEMI/Shock Stratified by Hospital Safety Net Status

Description	Safety Net Status	
	No	Yes
N	1024	202
Mean	0.0839	0.0864
Std Deviation	0.0144	0.0142
100% Max	0.1459	0.1280
99%	0.1234	0.1233
95%	0.1113	0.1115
90%	0.1027	0.1058
75% Q3	0.0929	0.0952
50% Median	0.0820	0.0844
25% Q1	0.0738	0.0762
10%	0.0666	0.0703
5%	0.0625	0.0666
1%	0.0553	0.0602
0% Min	0.0469	0.0539

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

N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Cardiovascular, Cardiovascular : Coronary Artery Disease (PCI)

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

Care Coordination, Safety, Safety : Complications

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

Elderly, Populations at Risk

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

https://www.ncdr.com/WebNCDR/docs/public-data-collection-documents/cathpci_v4_codersdictionary_4-4.pdf?sfvrsn=2

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** PCI_mortality_STEMI_Final-_With_NDI_Data_03Nov2017.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

No changes were made to the measure specification since the 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) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The outcome for this measure is all-cause death within 30 days following a PCI procedure in patients with STEMI or cardiogenic shock at the time of the procedure.

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

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

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). For the purpose of development and testing of the measure, we used a Medicare FFS population age 65 and over. We linked CathPCI registry with corresponding Medicare data and identified: a) in-hospital deaths using the discharge disposition indicator in the Standard Analytic File (SAF) and identified) post-discharge deaths using the Enrollment Database (EDB). For the purpose of maintenance, the measure used a cohort of patients whose vital status was determined from the National Death Index. This data sample reflects a more comprehensive data set including a broader age range (>18 years) and an all-payer model compared to the Medicare data set (>65 years) used for initial measure testing.

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

The target population for this measure includes inpatient and outpatient hospital stays with a PCI procedure for patients at least 18 years of age, with STEMI or cardiogenic shock at the time of procedure, including outpatient and observation stay patients who have undergone PCI but have not been admitted. It is unlikely that patients in this cohort would not be admitted to the hospital, but we keep this criterion to be consistent with the complementary non-STEMI, non-cardiogenic shock PCI cohort.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.) IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

The time window can be specified from one or more years. This measure was developed with Medicare claims and CathPCI Registry data from one calendar year.

The measure cohort is patients undergoing PCI who have STEMI or cardiogenic shock. STEMI or cardiogenic shock is defined as present in Version 4.4 of the CathPCI registry as follows:

Admissions with PCI are identified by field 5305 (PCI=yes);

STEMI or shock is identified by:

(1) Symptoms present on admission = ACS:STEMI (field 5000 = 6) with Time Period Symptom Onset to Admission within 24 hours (field 5005 = 5006, 5007, 5008) or Acute PCI = Yes (field 7035);

OR

(2) Cardiogenic shock = Yes (field 5060=1)

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

Hospital stays are excluded from the cohort if they meet any of the following criteria:

(1) PCIs that follow a prior PCI in the same admission (either at the same hospital or a PCI performed at another hospital prior to transfer).

This exclusion is applied in order to avoid assigning the death to two separate admissions.

(2) For patients with inconsistent or unknown vital status or other unreliable data (e.g. date of death precedes date of PCI);

(3) Subsequent PCIs within 30-days. The 30-day outcome period for patients with more than one PCI may overlap. In order to avoid attributing the same death to more than one PCI (i.e. double counting a single patient death), additional PCI procedures within 30 days of the death are not counted as new index procedures.

(4) PCIs for patients with more than 10 days between date of admission and date of PCI. Patients who have a PCI after having been in the hospital for a prolonged period of time are rare and represent a distinct population that likely has risk factors related to the hospitalization that are not well quantified in the registry.

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

Excluded hospital stays are identified as follows:

(1) PCIs that follow a prior PCI in the same admission or occur during a transfer-in admission (PCI to PCI). For the purposes of development we used Medicare data to define transfers as two admissions that occur within 1 day of each other and identified patients in this cohort who had a PCI during both admissions. This can also be identified in the registry data. (Note: For purposes of maintenance, we used NDI and CathPCI registry data)

(2) Patients with inconsistent or unknown vital status or other unreliable data (e.g. date of death precedes date of PCI). The specific data fields will depend on the data source used.

(3) Not the first hospital stay with a PCI in the 30 days prior to a patient death. These stays are identified by procedure date in the CathPCI Registry and death date in the vital status data source.

(4) PCIs for patients with more than 10 days between date of admission and date of PCI. We determine length of stay by subtracting the admission date from the procedure date in the CathPCI Registry

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

Results of this measure will not be stratified.

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

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

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

The measure score is calculated based on the following steps:

1. Patient cohort is identified based on the inclusion and exclusion criteria (see questions S.6, S.7, S.8, S.9, S.10);

2. Data elements for risk adjustment are collected using the first collected value, as detailed below;

3. Outcome is ascertained from an outside data source, such as the Medicare Enrollment Database (see questions S.4, S.5, S.6)

4. Measure score is calculated with aggregated data across all included sites, as described below.

Risk-adjustment variables

The measure is adjusted for the variables listed below:

1. Age (10 year increments)

2. Body Mass Index (5 kg/m^2 increments)

3. History of cerebrovascular disease

4. History of chronic lung disease

5. Glomerular Filtration Rate (GFR) (derived)

6. Previous PCI

7. Heart Failure - current status

8. Cardiogenic shock on admission

9. Symptom onset

10. Ejection Fraction percent (EF)

11. PCI status

12. Highest risk lesion – coronary artery segment category

13. Highest risk lesion: Society for Cardiovascular Angiography and Interventions (SCAI)

Measure Score Calculation

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 predicted hospital outcome (the numerator) 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 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 (better quality) and a higher ratio indicates higher-than-expected mortality (worse quality).

The predicted hospital outcome (the numerator) is calculated by regressing the risk factors and the hospital-specific intercept on the risk of mortality, multiplying the estimated regression coefficients by the patient characteristics in the hospital, transforming, then summing over all patients attributed to the hospital to get a value. The expected number of deaths (the denominator) is obtained by regressing the risk factors and a common intercept on the mortality outcome using all hospitals in our sample, multiplying the subsequent estimated regression coefficients by the patient characteristics observed in the hospital, transforming, and then summing over all patients in the hospital to get a value. To assess hospital performance in any reporting period, we re-estimate the model coefficients using the years of data in that period.

Please see attachments for more details on the calculation algorithm and the value sets for the risk-adjustment variables.

References:

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22 (2): 206-226.

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

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

N/A. This measure is not based on a sample or survey. Data from all hospitals and all PCI procedures would be included in the process of re-estimating model variables. For public reporting, minimum sample size has not been determined.

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

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

N/A. This measure is not based on a sample or survey.

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.

Claims, Other, Registry Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*
IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

Data sources:

NCDR CatchPCI Registry

Vital Status Source: National Death Index, Death Masterfile, Medicare enrollment database, or equivalent

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

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

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

Facility, Other

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

Inpatient/Hospital

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

0536_NQF_testing_attachment_20171108.pdf

2.1 For maintenance of endorsement

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

No

2.2 For maintenance of endorsement

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

No

2.3 For maintenance of endorsement

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

Yes - Updated information is included

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

Measure Number (*if previously endorsed*)**:** 0536
Measure Title: 30-day all-cause risk-standardized mortality rate following Percutaneous Coronary Intervention (PCI) for patients with ST segment elevation myocardial infarction (STEMI) or cardiogenic shock
Date of Submission: Click here to enter a date11/8/2017

Type of Measure:

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

Instructions

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

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

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

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

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

AND

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

Notes

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

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

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

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

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

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

1. DATA/SAMPLE USED FOR <u>ALL</u> 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. <u>If there are differences by aspect of testing,</u>(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.17*)	Measure Tested with Data From:
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☒ registry	☒ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Source of vital status (e.g. National death index)	☒ other: Source of vital status (e.g. National death index)

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*).

Medicare Part A claims, National Cardiovascular Data Registry (NCDR) CathPCI Registry,

Medicare Enrollment Database

We linked CathPCI Registry and Medicare data and identified in-hospital deaths using the discharge disposition indicator in the Standard Analytic File (SAF) and identified post-discharge deaths using the Enrollment Database (EDB)

1.3. What are the dates of the data used in testing? 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.20*)	Measure Tested at Level of:
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ 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*)

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 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:

Measure reliability and validity dataset

The measure reliability and validity dataset linked the CathPCI and Medicare Part A claims data from 2010-2011. It included 48,339 admissions to 1,182 hospitals with 24,170 admissions to 1,167 hospitals in one randomly selected sample and 24,169 admissions to 1,160 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 360 hospitals and the second hospital contained 360 hospitals. The linked dataset was used for:

- Data element reliability testing (Section 2a2)
- Measure score validity testing (Section 2b2)
- Measure exclusions testing (Section 2b3)

Data validity (Section 2b2)

We used admissions of patients discharged from January through December 2005.

Risk adjustment dataset (Section 2b4)

We use admissions with PCI in the merged data from 2006. The development sample consisted of 15,123 admissions at 602 hospitals in the STEMI or shock cohort.

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

Social risk factors were not used in this risk model for the following reasons. First, as a detailed clinical registry used for quality assessment and improvement, there are not prospective interviews with patients to obtain patient-reported data. Second, the effect of social risk factors may be at either the patient- or the hospital-level. For example, patients with social risk factors (i.e., low income, lack of education, etc.) may have an increased risk of mortality because these patients may have an individual higher risk (patient-level effect) or because patients with social risk factors are more often admitted to hospitals with higher overall mortality rates (hospital-level effect). It is important to note, however, that even in the presence of a significant patient-level effect and absence of a significant hospital-level effect, the increased risk could be partly or entirely due to the quality of care patients receive in the hospital. For example, biased or differential care provided within a hospital to low-income patients as compared to high-income patients would exert its impact at the level of individual patients, and therefore be a patient-level effect. Third, while it may be true that worse social risk factors might be associated with more severe illness at the time of presentation, we had direct access to detailed clinical variables describing the severity of illness and feel that incorporating such factors (e.g. cardiogenic shock, ejection fraction, PCI status, cardiac arrest, highest risk legion, etc.) is a much more accurate means of stratifying risk. Accordingly, we feel that in this model of 30-day AMI mortality for STEMI/Shock patients, given the rich clinical data available through the NCDR CathPCI Registry and linkage to National Death Index data, that social risk factors, which are not readily available, would not likely contribute much improvement to this particular risk model.

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

See Section 2b2 for validity testing of data elements

Measure Score Reliability

To assess reliability of the measure, we examined the extent to which assessments of a hospital using different but randomly selected subsets of patients in the same time period produced similar measures of hospital performance. That is, we took 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 calculated the agreement of the two resulting performance measures across hospitals.

For test-retest reliability of the measure in Medicare FFS patients aged 65 and older, we combined index admissions from two years (2010 and 2011) into a single dataset, randomly sampled half of patients within each hospital, calculated the measure for each hospital, and repeated the calculation using the second half. Thus, we measured each hospital 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 reliable. As a metric of agreement we calculated the intra-class correlation coefficient and assessed the values according to conventional standards.

Specifically, we used a combined 2010-2011 sample that had been linked with Medicare FFS claims data, and randomly split it into two approximately equal subsets of patients. We then calculated the RSMR for each hospital for each sample. The agreement of the two RSMRs was quantified for hospitals in each sample using the intra-class correlation. 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. Of note, because our final measure is derived using hierarchical logistic regression, a known property of hierarchical logistic regression models is that smaller 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 PCI population, but we tested it only in the subset of Medicare FFS patients for whom information about vital status was available. This reduced the cohort available for testing by approximately 40%.

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*)

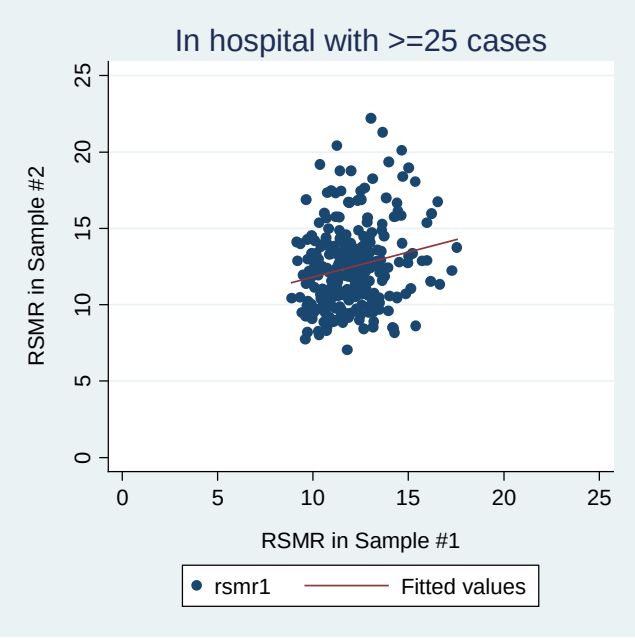
Measure Score Reliability

We calculated the correlation of the RSMR from our final model in two different samples.

Table 1. Overall mortality rate (OMR) and risk-standardized mortality rate (RSMR) in the split samples; 2010-2011.

Description	First Half of the Data			Second Half of the Data		
		Weighted by Hospital Volume			Weighted by Hospital Volume	
	Volume	OMR	RSMR	Volume	OMR	RSMR
N	1,167	24,170	24,170	1,160	24,169	24,169
Mean	20.71	0.1230	0.1245	20.84	0.1209	0.1211
Std Deviation	17.48	0.0799	0.0249	17.19	0.0775	0.0152
100% Max	136	1.0000	0.2274	141	1.0000	0.1754
99%	90	0.3750	0.2041	84	0.3333	0.1663
95%	52	0.2667	0.1700	55	0.2500	0.1498
90%	42	0.2195	0.1575	43	0.2174	0.1413
75% Q3	28	0.1667	0.1374	28	0.1579	0.1299
50% Median	17	0.1154	0.1216	17	0.1111	0.1189
25% Q1	9	0.0667	0.1067	9	0.0714	0.1103
10%	3	0.0303	0.0952	3	0.0357	0.1037
5%	2	0.0000	0.0905	2	0.0000	0.0993
1%	1	0.0000	0.0799	1	0.0000	0.0940
0% Min	1	0.0000	0.0703	1	0.0000	0.0888

Figure 1. Correlation between Hospital Risk-Standardized Mortality Rates in Split Samples; 2010-2011.



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 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. Accordingly, our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients in the same time period produce similar measures of hospital performance. The agreement between the two RSMRs (0.122), which according to conventional interpretation is "slight," likely reflects the relatively low number of cases included in the cohort as outlined above (Landis JR et al. 2013). Nevertheless, the reliability of the measure should be assessed using larger split samples when available. Based on our experience with similar measures using split samples, using 4 years (and volume equivalent to 2 years) would result in higher intra-class correlation coefficient.

References

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. Mar 1977;33(1):159-174.

2b1. VALIDITY TESTING

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

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

☐ **Performance measure score**

☒ **Empirical validity testing**

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

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

Data Element Validity

Data element validity testing was done on the specified measure by comparing with variables in the ACC audit program. The NCDR CathPCI Registry has an established Data Quality Program 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 PCI status. The DAP includes an appeals process for hospitals to dispute the audit findings. The NCDR DAP was accepted by the National Quality Forum as part of its endorsement of the CathPCI Registry's in-hospital risk-adjusted mortality measure.

10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) Code Selection

In 2012, we used the General Equivalence Mapping (GEM) crosswalk between ICD-9-CM and ICD-10-CM/PCS to create specifications for the measure in ICD-10-CM/PCS. Our process for mapping procedural codes in the measures to ICD-10-CM consisted of a detailed clinical review, including manual review of related ICD-10-CM codes to determine that all appropriate codes are included, rather than relying exclusively on the GEM. To conduct the crosswalk, we created a database to effectively use the mapping tables provided by CMS. We then compiled a list of ICD-9-CM codes that define PCI during hospitalization. Measure developers used these ICD-9-CM codes to build queries to extract the GEM results from the mapping table in the database. We then applied those ICD-10-CM codes to the ICD-10-CM to ICD-9-CM mapping table to see if the reverse query produced ICD-9-CM codes that were not in the original measure specifications.

Our clinicians reviewed these results in detail and determined that many ICD-10-CM codes that should be included in our cohort were not being captured by the GEMs. We confirmed this by consulting the ICD-10-CM draft procedural codebook and identifying the ICD-10-CM codes that our clinicians felt should be included in our cohort. The GEMs identified 16 ICD-10-CM codes for our PCI mortality cohort, while clinician review of the ICD-10-CM draft codebook resulted in 48 ICD-10-CM codes.

Further details also are located in the attached Appendix.

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

Data Element Validity

In the audit that assessed cases submitted in 2005, the median agreement between submitted and audited values was 92%. There was consistency across sites, with agreement in the lowest and highest deciles of hospitals ranging from 90% to 95%.

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

Data Element Validity

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.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section* 2b4

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

Exclusions were those determined by expert input to be clinically relevant, required in order to assess the outcome, or needed for calculation of the measure. To ascertain the impact of the exclusions on the cohort, we examined proportions of the total cohort excluded for each exclusion criterion

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

Table 2. Exclusions from the target population for the 2010, 2011, and the combined 2010-2011 dataset.

Exclusions	2010		2011		2010-2011
	Patient Stay	Hospitals	Patient Stay	Hospitals	Patient Stay
	# (%)	# (%)	# (%)	# (%)	# (%)
Initial Sample	**199,853**	**1,095**	**195,812**	**1,185**	**395,665**
Not Medicare patient on admission	43, 669 (21.85)	3 (0.27)	44,840 (22.90)	1 (0.08)	88509 (22.37)
Remaining	**156184**	**1,092**	**150,972**	**1,184**	**307,156**
Not the first claim in the same claim bundle*	3 (0.00)	0 (0.00)	8 (0.01)	0 (0.00)	11 (0.00)
Remaining	**156181**	**1,092**	**150,964**	**1,184**	**307,145**
Get the procedure more than 10 days after admission	1,074 (0.69)	0 (0.00)	1,212 (0.80)	0 (0.00)	2286 (0.74)
Remaining	**155,107**	**1,092**	**149,752**	**1,184**	**304,859**
Transferred in (PCI to PCI)	186 (0.12)	0 (0.00)	204 (0.14)	(0.00)	390 (0.13)
Remaining	**154,921**	**1,092**	**149,548**	**1,184**	**304,469**
Unknown death	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Remaining	**154,921**	**1,092**	**149,548**	**1,184**	**304,469**
Duplicate death	65 (0.04)	0 (0.00)	77 (0.05)	0 (0.00)	142 (0.05)
Remaining	**154,856**	**1,092**	**149,471**	**1,184**	**304,327**
AMA	215 (0.14)	0 (0.00)	212 (0.14)	0 (0.00)	427 (0.14)
Remaining	**154,641**	**1,092**	**149,259**	**1,184**	**303,900**
Not with STEMI/Shock	130,942 (84.67)	20 (1.83)	124619 (83.49)	28 (2.36)	255561 (84.09)
Study Sample	**23,699**	**1,072**	**24,640**	**1,156**	**48,339**
Death within 30-days from procedure	2,804 (11.83)		3090 (12.54)		5894 (12.19)
In-Hospital death	2,241 (9.5)		2439 (9.9)		4680 (9.68)

* Defined as two or more claims in which the admission date of the current claim is before or the same as the discharge date of its previous claim. When this happens, the information at discharge of the first claim are replaced by the information at discharge of the last claim.

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis.*

*Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The decision to exclude patients discharged AMA is based on clinical judgment to make the measure fair and is unlikely to distort the results given the very low frequency. Excluding patients transferring into a hospital does not actually exclude acute episodes from the measure, but considers the hospital that initially admits the patient as the one accountable for the outcome, avoiding double counting and clarifying accountability. The exclusion of unreliable data is necessary for valid calculation of the measure. Excluding PCIs that follow a prior PCI in the same admission or during a transfer-in is applied in order to avoid assigning the death to two separate admissions. The decision to exclude subsequent PCIs within 30 days of death is necessary to avoid attributing the same death to more than one PCI. Lastly, patients who get the procedure more than 10 days after admission have a PCI after many days of hospitalization are rare and represent a distinct population that likely has risk factors related to the hospitalization and not well quantified in the registry.

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

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 13 risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

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

N/A

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

The goal of risk adjustment is to account for different patient demographic and clinical characteristics at the time of admission (hospital case mix), enabling interpretation of any identified differences in quality. Conditions that may represent adverse outcomes due to care received during the index hospital stay are not included in the risk-adjustment model. We sought to develop a model that included key variables that were clinically relevant and based on strong association with 30-day mortality.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

To create a model with increased usability while retaining excellent model performance, we tested the performance of the model without those variables considered to be questionably feasible. To select candidate variables, a team of

clinicians reviewed all variables in the NCDR CathPCI 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). We did not consider as candidate variables those that we would not want to adjust for in a quality measure, such as potential complications, certain patient demographics (e.g., race, socioeconomic status), and patients‟ admission path (e.g., admitted from, or discharged to, a skilled nursing facility [SNF]). Variables were also considered ineligible if they were particularly vulnerable to gaming or were deemed to lack clinical relevance. Based on careful review by a team of clinicians and further informed by a review of the literature, a total of 26 variables were determined to be appropriate for consideration as candidate variables. Our set of candidate variables included two "demographic" variables (age and gender), 15 "history and risk factor" variables, four "cardiac status" variables, one "cath lab visit" variable and four "PCI procedure" variables. The final risk-adjustment model for the STEMI or shock cohort included 13 variables:

1) Age

2) Body mass index (BMI)

3) Cerebrovascular disease

4) Chronic lung disease

5) Glomerular filtration rate (GFR)

6) Previous PCI

7) Congestive heart failure (CHF) status

8) Cardiogenic shock

9) Symptoms present on admission

10) Ejection fraction percentage

11) PCI status

12) Highest risk lesion – segment category

13) Highest risk lesion – SCAI lesion class

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

N/A

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model <u>or</u> stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Several variables required particular consideration. First, in the current version of the CathPCI registry, participants are instructed to use New York Heart Association (NYHA) classification to capture symptom severity for both heart failure and angina. Accordingly, the resulting variable is a hybrid which may dilute the prognostic importance usually associated with NYHA class. Second, variables such as PCI status and cardiogenic shock impart important prognostic information but are vulnerable to systematic misclassification. This is relevant to efforts to publicly report 30-day PCI mortality in that several key variables (e.g., cardiogenic shock and PCI status) may be consistently coded differently across sites. For example, although the CathPCI data dictionary provides detailed definitions of PCI status (http://www.ncdr.com/WebNCDR/ELEMENTS.ASPX), sites may differ in their interpretation of these definitions such that a patient considered an emergent PCI at hospital A may be considered an urgent PCI at hospital B. If differences in coding occur with sufficient frequency, the risk-standardized mortality rate for hospital A might appear lower than hospital B, even if their case mixes and outcomes were otherwise identical.

To examine this issue, we compared the frequency of different PCI status categories at hospitals with risk adjusted mortality rates that were above and below the median using the STEMI or shock cohort. We found that rates of cardiogenic shock were comparable, but that hospitals with below average risk-standardized mortality had modestly higher rates of emergency and salvage PCI (76.7% and 1.4%), compared with hospitals with above average risk-

standardized mortality (72.3% and 1.2%). We cannot determine whether these differences accurately reflect differences in case mix or are due to systematic differences in coding. Nevertheless, these results highlight the need to further ensure data accuracy.

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%). There were three continuous variables with significant numbers of missing values: body mass index (BMI), glomerular filtration rate (GFR), and left ventricular ejection fraction (LVEF). For BMI, we stratified by gender and imputed the missing values to the median of the corresponding groups. For GFR, we stratified patients into five categories: <30, 31-60, 61-90, >90, and missing. For LVEF, we stratified patients into four categories- <30%, 31-45%, >45%, and missing.

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.

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

If stratified, skip to 2b3.9

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

We computed 6 summary statistics for assessing model performance: over-fitting indices, percentage of variation explained by the risk factors, predictive ability, area under the receiver operating characteristic (ROC) curve, distribution of residuals, and model chi-square.

The development model has excellent discrimination, calibration, and fit. The patient-level mortality rate ranges from 1.4% in the lowest predicted decile to 40.3% in the highest predicted decile, a range of 38.9%. The area under the ROC curve is 0.825.

The discrimination and the explained variation of the model at the patient-level are consistent with those of published PCI in-hospital mortality models (Yale-CORE 2008). The ROC is modestly lower than that of previously published models due to several factors. First, we stratified the entire population of PCI patients into two populations based on the presence or absence of two prognostically important variables: STEMI and cardiogenic shock. Second, we excluded covariates such as potential complications, certain patient demographics (e.g., race), and patients‟ admission path (e.g., outpatient, emergency department, transfers-in from other facilities (non-acute care or acute care). These characteristics may be associated with mortality and thus could increase the model performance to predict patient mortality. However, these variables may be related to quality or supply factors that should not be included in an adjustment that seeks to control for patient clinical characteristics. Thus, the choice was to focus on adjustment for clinical differences in the populations among hospitals.

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

Table 3. Model Performance: Calibration Results Based on the Logistic Regression Model

Indices	2011 Sample	2010 Sample
Number of Admissions	24,640	23,699
Calibration		
γ0, γ1	-	-0.088, 0.997
ROC	0.827	0.831
Residuals Lack of Fit (Pearson Residual Fall %)		
<-2	0.175	0.186
[-2, 0)	87.285	87.983

Indices	2011 Sample	2010 Sample
[0, 2)	7.171	6.675
[2+	5.369	5.156

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

Table 4. RSMR Model Performance for STEMI or Shock Cohort

Indices	Development Sample (2010)	Validation Sample (2011)	Merged Sample (2010-2011)
Number of hospitals	1,072	1,156	1,182
Number of admissions	23,699	24,640	48,339
RSMR			
100% Max	0.2077	0.2275	0.1983
99%	0.1747	0.1906	0.1813
95%	0.1578	0.1661	0.1616
90%	0.1452	0.1542	0.1496
75%	0.1297	0.1378	0.1331
50% Median	0.1159	0.1249	0.1201
25%	0.1060	0.1113	0.1094
10%	0.0977	0.1007	0.0992
5%	0.0936	0.0953	0.0935
1%	0.0834	0.0841	0.0848
0% Min	0.0751	0.0741	0.0778

2b3.9. Results of Risk Stratification Analysis:

N/A

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

The C-statistic of 0.825 indicates excellent model discrimination. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model. The risk decile plot shows excellent discrimination of the model and good predictive ability.

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

N/A

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

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

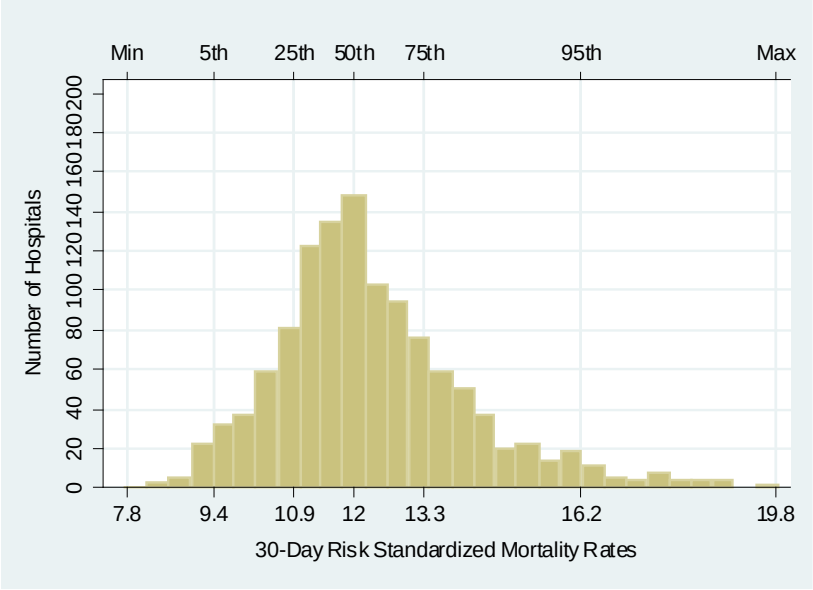
For the currently publicly reported measures of hospital outcomes, including the PCI readmission measure, CMS estimates an interval estimate for each risk-standardized rate to characterize the amount of uncertainty associated with the rate. It then compares the interval estimate to the national crude 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 PCI). However, the decision to publicly report this PCI mortality measure and the approach to discriminating performance has not been determined.

We assessed variation in RSMRs among hospitals by examining the distribution of the hospital RSMRs and plotting the histogram of the hospital RSMRs.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

In the 2010-2011 sample, the mean hospital RSMR for the STEMI or shock cohort was 12.3%%, with a range of 7.8% to 19.8%. The interquartile range was 10.9% to 13.3%.

Figure 2. Distribution of risk-standardized mortality rates (RSMRs); 2010-2011 combined sample.



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

The variation in rates suggests there are meaningful differences across hospitals in the 30-day risk-standardized mortality after PCI in the STEMI or shock cohort.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

N/A

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

N/A

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

N/A

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

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

We examined rates of missing data for all candidate variables and examined histograms of the frequency of missingness by hospital.

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

Overall the percentage of missing values for all categorical variables was very small (<1%). There were three continuous variables with significant numbers of missing values: body mass index (BMI), glomerular filtration rate (GFR), and left ventricular ejection fraction (LVEF). The frequency of missingness by hospital appeared to be evenly distributed across hospitals. Model performance and estimates of hospital RSMR were not significantly different when repeated excluding cases with missing data. The fact that the data was missing did not appear to be at random in that patients with missing data regarding GFR, and LVEF were at higher risk of death than those without missing data. Accordingly we created a dummy variable to capture that information.

For categorical variables with missing values, the value from the reference group was added. For BMI, we stratified by gender and imputed the missing values to the median of the corresponding groups. For GFR, we stratified patients into five categories: <30, 31-60, 61-90, >90, and missing. For LVEF, we stratified patients into four categories- <30%, 31-45%, >45%, and missing.

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

As noted above, model performance was comparable when we included or excluded cases with 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.

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), Other

If other: The outcome will be determined from an administrative database such as the National Death Index.

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic health records (EHRs)

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

ACC is in the process of developing a common data dictionary mapped to coded terminology standards with the intent of improving interoperability with EHRs and potentially creation of emeasures.

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

Attachment:

3c. Data Collection Strategy

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

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

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

Implementation of this measure requires matching of patient data to external data source to determine the outcome endpoint (Death 30 days after PCI). This has resulted in several implementation challenges

- Data Availability: ACC is not able to use CMS data as a source for this measure as it is not being used for research purposes (CMS ResDAC path) and we do not have other payor data after the hospital visit (Qualified Entity requirement) and do not fit either path to receive CMS Data. We have had to change implementation strategy, rework our models and match NCDR records to CDC National Death Index (NDI) data.
- Patient Confidentiality: CDC NDI requires direct patient identifiers in order to meet the minimum criteria for matching. Roughly 15% of submitting NCDR sites (based on 2017Q2 CathPCI data) do not submit direct patient identifiers to the registry and are therefore ineligible for NDI matching and cannot participate in this measure.
- Data Cost: CDC NDI charges for matching of data to NDI. This results in a $100,000 cost to ACC per year to report this measure for the CathPCI registry.
- Data Timeliness: CDC NDI is released on a yearly basis, roughly one year after the calendar year of death along with processing time of the matching process and report generation and the most contemporary data available is over 18 months old. In addition, calculating 30 Day mortality for December of the calendar year requires waiting for an extra 12 months (30 months total) in order to get death certificates for January of the following year.

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)*.

This measure was developed and designed to be used across other organizations and by other measure implementers. The fee and licensing information included below is specific to NCDR program requirements:

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 2017 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 CathPCI 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.

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting	Quality Improvement (external benchmarking to organizations) National Cardiovascular Data Registry https://www.ncdr.com/webncdr/cathpci/home/datacollection

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

N/A

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

This measure is designed for use in public reporting, but it is currently not in use. See 4a1.3 for rationale and plan for public reporting. ACC plans to include this measure in NCDR's public reporting program in the future.

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

Update to credible plan (11/8/17):

We moved forward with implementing the 30-day risk adjusted mortality measures in the CathPCI registry for the ´Quarter 3, 2017 30-Day mortality outcomes report´ which included data from 2011 to 2014. However, ACC held off on public reporting since we are also in the process of updating the CathPCI registry to version 5. The new registry version includes elements to assess out-of-hospital cardiac arrest, which has been identified in the literature as a risk factor that should be considered in mortality modeling(1,2). Additionally, when preparing the public reporting metric for in-hospital mortality (#0133) and 30-day mortality (#0536), we found that the measures were not harmonized in structure (i.e. the 30-day measure is a hierarchical model whereas the in-hospital measure is not). As such, these measures could not be rolled up together to create an appropriate composite view of mortality. We plan to modify the in-hospital mortality model to a hierarchical structure when we expand to take advantage of the additional elements in version 5 of CathPCI registry, particularly cardiac arrest, rather than sequencing a number of major revisions in a relatively short time period for hospitals. In order to avoid unintended negative consequences, ACC has made the decision to put a hold on public reporting until the cardiac arrest elements can be considered for modeling and the inpatient and 30-day PCI mortality models can be structurally harmonized. In addition, for purposes of public reporting this measure will also always be paired with (#0535) 30-day all-cause risk-standardized mortality rate following percutaneous coronary intervention (PCI) for patients without ST segment elevation myocardial infarction (STEMI) and without cardiogenic shock.

Citation:

[1] Peberdy, M.A., Donnino, M.W., Callaway, C.W., et al. Impact of Percutaneous Coronary Intervention Performance Reporting on Cardiac Resuscitation Centers: A Scientific Statement From the American Heart Association. Circulation. 2013;128:762-773; originally published online July 15, 2013; doi: 10.1161/CIR.0b013e3182a15cd2

[2] Camuglia, A.C., Randhawa, V.K., Lavi, S., et al. Cardiac catheterization is associated with superior outcomes for survivors of out of hospital cardiac arrest: Review and meta-analysis. Elsevier: Resuscitation 85 (2014) 1533–1540 . www.elsevier.com/locate/resuscitation

NCDR Public Reporting Background:

ACC's National Cardiovascular Data Registry (NCDR) Voluntary Hospital Public Reporting Program: The ACC currently runs a program to give hospitals the opportunity to voluntarily publicly report their measure results based on data from the National Cardiovascular Data Registry (NCDR). Hospitals that choose to participate have their results displayed on ACC's CardioSmart. Currently Hospitals can report on five measures from the CathPCI Registry and five measures from the ICD Registry. Of these publicly reporting measures, five are NQF-endorsed:

- NQF # 1522: Use of a medicine in the ACEi or ARB class to improve heart function after ICD implant in patients with less than normal heart function.
- NQF # 1528: Use of a beta-blocker medication after ICD implant in patients with a previous heart attack.
- NQF #1529: Use of a beta-blocker medication after ICD implant in patients with less than normal heart function.
- NQF #0965: Use of all recommended medications (ACEI or ARB and beta-blocker) to improve heart function and blood pressure after ICD implant.
- NQF # 0964: Therapy with aspirin, P2Y12 inhibitor, and statin at discharge following PCI in eligible patients (composite measure)

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

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

Performance results are distributed to all CathPCI registry participants as part of benchmark reports, which provide a detailed analysis of an institution´s individual performance in comparison to the entire registry population from participating hospitals across the nation. Reports include an executive summary dashboard, at-a-glance assessments, and patient level drill-downs. Registry participants also have access to an outcome report companion guide which provides common definitions and detailed metric specifications to assist with interpretation of performance rates.

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

The majority of the required data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions can manually report using a free web-based tool or automate the reporting by using certified software developed by third-party vendors. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

There are a number of methods used to educate and provide general support to registry participants. This includes the following:

- Registry Site Manager Calls are available for all NCDR participants. RSM calls are provided as a source of communication between NCDR and participants to provide a live chat Q and A session on a continuous basis.
- New User Calls are available for NCDR participants, and are intended for assisting new users with their questions.
- NCDR Annual Conference

The NCDR Annual Conference is a well-attended and energetic two-day program at which participants from across the country come together to hear about new NCDR and registry-specific updates. During informative general sessions, attendees can learn about topics such as transcatheter therapies, the NCDR dashboard, risk models, data quality and validation, and value-based purchasing. Attendees also receive registry updates and participate in advanced case studies covering such topics as Appropriate Use Criteria and outcomes report interpretation.

- Release notes (for outcomes reports)
- Clinical Support

The NCDR Product Support and Clinical Quality Consultant Teams are available to assist participating sites with questions Monday through Friday, 9:00 a.m. - 5:00 p.m. ET.

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

Describe how feedback was obtained.

Feedback is typically obtained through monthly registry site manager monthly calls, ad hoc phone calls tracked with salesforce software, and during registry –specific break-out sessions at the NCDR's annual meeting. Registry Steering Committee members may also provide feedback during regularly scheduled calls.

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

While the 30-day mortality measure was implemented for the first time in the CathPCI registry in quarter 3, 2017, the registry participants appear to be very interested in this measure. However, since it was implemented relatively recently, there have been no major issues or other feedback received from registry participants with respect to collecting data for this particular metric.

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

No other feedback was received from other users.

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

N/A (Measure was not modified since last endorsement)

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

The performance data used and described in 1b reflects a different cohort of data from when the measure was last endorsed. We previously analyzed CMS and CathPCI registry data from 2010 to 2011, however, for this endorsement period had access to the National Death Index (NDI) data from 2011-2014. NDI data is more comprehensive and allowed for the risk model to be applied to all-payers and a wider age range of patients (>18) compared to CMS data (>65). Based on the differences in cohorts of data analyzed (CMS vs NDI), we are unable to comment on or draw conclusions from risk adjusted performance trends over time. However, the unadjusted 30-day mortality rate was 7.9% in 2011-12, then it increased slightly to 8.3% in 2012-2013 and then decreased to 7.4% in 2013-14.

4b2. Unintended Consequences

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

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

Studies suggest that public reporting of the outcomes of cardiovascular procedures may have unintended consequences. Joynt and colleagues compared the characteristics and outcomes of patients undergoing PCI in states with (MA, NY, PA) and regional states without (CT, DE, ME, MD, NH, RI, VT) public reporting and found that patients with acute MI were less likely to receive PCI in public reporting states than in non-public reporting states. There were no differences in overall 30-day mortality rates among acute MI patients in reporting versus non-reporting states. Determining the underlying causes and appropriateness of these differences is impossible, but there is concern that physicians in states that publicly report PCI outcomes would either refer high risk cases to states without public reporting or avoid such cases altogether. Implementing a national measure of PCI outcomes would avoid the former problem in that public reporting would be consistent across states.

Nevertheless, this measure will continue to require close attention to the possibility that high risk patients are not receiving PCI when clinically indicated. The measure is, however, complementary to the previously approved measures for 30-day mortality of AMI and heart failure patients in that inappropriate avoidance of high risk PCI cases may have a detrimental effect on hospitals' performance on these other measures of cardiovascular outcomes. However, it is important to note that this measure has not undergone public reporting to date, thus the unintended consequences are speculative.

Measure implementation will require close attention to data quality. Potential solutions include a) detailed chart audits, b) close attention to variances in case mix and c) review of some or all cases coded as cardiogenic shock or a salvage PCI.

Joynt, K. E., Blumenthal, D. M., Orav, E. J., Resnic, F. S., & Jha, A. K. (2012). Association of Public Reporting for Percutaneous Coronary Intervention with Utilization and Outcomes among Medicare beneficiaries with Acute Myocardial Infarction. JAMA?: The Journal of the American Medical Association, 308(14), 1460–1468. http://doi.org/10.1001/jama.2012.12922

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

N/A - there were no unexpected benefits noted for 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)

0229 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older

0230 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older

0535 : 30-day all-cause risk-standardized mortality rate following percutaneous coronary intervention (PCI) for patients without ST segment elevation myocardial infarction (STEMI) and without cardiogenic shock

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

NQF # 0535 - 30-day all-cause risk-standardized mortality rate following percutaneous coronary intervention (PCI) for patients without ST segment elevation myocardial infarction (STEMI) and without cardiogenic shock

NQF # 0230 - Acute Myocardial Infarction 30-day Mortality

NQF # 0229 - Heart Failure 30-day Mortality

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

N/A

5b. Competing Measures

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

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

This measure is most similar to the 30-day all-cause risk-standardized mortality rate following percutaneous coronary intervention (PCI) for patients without ST segment elevation myocardial infarction (STEMI) and without cardiogenic shock. Its additive value stems from the target population of STEMI and/or shock patients.

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: PCI_Mortality_STEMI_Appendix_Attachment-636426313699895942.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American College of Cardiology

Co.2 Point of Contact: Kim, Lavin, comment@acc.org, 202-375-6448-

Co.3 Measure Developer if different from Measure Steward: American College of Cardiology

Co.4 Point of Contact: Esteban, Perla, eperla@acc.org, 202-375-6499-

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 measure developer, Yale New Haven Health Servicec Corporation Center for Outcomes Research and Evaluation (YNHHSC/CORE) obtained expert and stakeholder input on the two measures through two mechanisms. First, the team has held regular conference calls with a Working Group of YNHHSC/CORE and American College of Cardiology (ACC)/National Cardiovascular Data Registry (NCDR) experts in cardiovascular registries and in the outcomes measure field. Second, YNHHSC/CORE sought and considered the input of an American College of Cardiology Foundation (ACCF) designated Task Force.

Working Group

Ralph Brindis, M.D., M.P.H., F.A.C.C.

Regional Senior Advisor for Cardiovascular Disease, Northern California Kaiser Permanente; Clinical Professor of Medicine, UCSF, Oakland, CA; Chief Medical Officer and Chairman, Management Board, National Cardiovascular Data Registry

Barbara Christensen, R.N., M.H.A.

Senior Director, Registry Services, American College of Cardiology

Jeptha Curtis, M.D.

Assistant Professor of Medicine, Department of Internal Medicine (Cardiovascular Disease), Yale University

Elizabeth Drye, M.D., S.M.

Research Project Director, Yale/Yale-New Haven Hospital Center for Outcomes Research and Evaluation

Susan Fitzgerald, R.N., M.B.A.

Associate Director, Registry Development, American College of Cardiology

Lori Geary, M.P.H.

Research Project Coordinator, Yale/Yale-New Haven Hospital Center for Outcomes Research and Evaluation

Amy Heller, Ph.D., M.P.H.

Associate Director, Quality Products, American College of Cardiology

Tony Hermann, R.N., M.B.A., C.P.H.Q.

Associate Director, CathPCI Registry, American College of Cardiology

Kathleen Hewitt, R.N., M.S.N., C.P.H.Q.

Associate Vice President, American College of Cardiology

Harlan Krumholz, M.D., M. Sc., F.A.C.C.

Director, Yale Center for Outcomes Research and Evaluation; Representative, NCDR analytic center; Ex-officio to Task Force

Kristi Mitchell, M.P.H.

Senior Director, Research, Development and Quality Products, American College of Cardiology

Eric Peterson, M.D., M.P.H., F.A.C.C.

Professor of Medicine, Duke University; Director, Cardiovascular Outcomes, Duke Clinical Research Institute, Chapel Hill, NC; Member, NCDR Science Oversight Committee/ Representative, NCDR Analytic Center

John Rumsfeld, M.D., Ph.D., F.A.C.C.

Associate Professor of Medicine, University of Colorado; Clinical Coordinator, VA Ischemic

Lara Slattery, M.H.S.

Associate Director, Research, Development, and Quality Products Department – Registries, Products, and Publishing Division, American College of Cardiology

John Spertus, M.D., M.P.H., F.A.C.C.

Director of Cardiovascular Education and Outcomes Research, Mid America Heart Institute, Kansas City, MO; Member, NCDR Science Oversight Committee/Representative, NCDR analytic center; Chair, American College of Cardiology Foundation Task Force on Public Reporting of Hospital-Level Outcomes Measures

Yongfei Wang, M.S.

Senior Research Analyst, Yale/Yale-New Haven Hospital Center for Outcomes Research and Evaluation

William Weintraub, M.D., F.A.C.C.

Chair, CathPCI Registry Steering Committee; Section Chief, Cardiology, Christiana Care Health Services, Inc., Newark DE

Al Woodward, Ph.D., M.B.A.

Director, Research Services, American College of Cardiology

Task Force

Five Task Force members also serve as members of the Working Group, including:

Ralph G. Brindis, M.D., M.P.H., F.A.C.C.

Harlan Krumholz, M.D., M. Sc., F.A.C.C.

Eric Peterson, M.D., M.P.H., F.A.C.C.

John Rumsfeld, M.D., Ph.D., F.A.C.C.

John Spertus, M.D., M.P.H., F.A.C.C.

Other Task Force members are:

John Brush, M.D., F.A.C.C.

Cardiology Consultants LLC, Norfolk, VA; Chair, Quality Strategic Directions Committee

Vincent J. Bufalino, M.D., F.A.C.C.

Midwest Heart Specialists, Naperville, IL; Co-Chair, ACC Advocacy Committee

Gregory Dehmer, M.D., F.A.C.C.

Professor of Medicine, Texas A&M College of Medicine, Temple, TX;

Representative, The Society for Cardiovascular Angiography and Interventions

James Dove, M.D., F.A.C.C.

President, American College of Cardiology

President Emeritus, Prairie Cardiovascular Consultants, Ltd., Springfield, IL; President, ACC/ACCF Board of Trustees

Stephen C. Hammill, M.D., F.H.R.S.

Professor of Medicine, Mayo Clinic College of Medicine, Rochester, MN; Representative, Heart Rhythm Society

Frank E Harrell Jr., PhD

Professor of Biostatistics; Department Chair, Vanderbilt University School of Medicine- Department of Biostatistics, Nashville, TN

Barry K. Lewis, D.O., F.A.C.C.

Consultants in Cardiology, P.C., Farmington Hills, MI; Member, Advocacy Committee

William R. Lewis, M.D., F.A.C.C.

Metro Health Medical Center, Cleveland, OH; ACC Ohio Chapter Governor/ACC Board of Governors

Fred Masoudi, M.D., M.S.P.H., F.A.C.C.

Denver Health Medical Center, Denver, CO; Chair, ACC/AHA Task Force on Performance Measures

Andrea M. Russo, M.D. F.A.C.C.

University of Pennsylvania Health System, Philadelphia, PA; Representative, Heart Rhythm Society

Bonnie H. Weiner, M.D., F.S.C.A.I., F.A.C.C.

Professor of Medicine; Interim Chair Cardiovascular Medicine, St. Vincent Hospital at Worcester Medical Center, Worchester, MA; Representative, The Society for Cardiovascular Angiography and Interventions

Stuart Winston, D.O., F.A.C.C.

Michigan Heart, P. C., Ann Arbor, MI; ACC Michigan Chapter Governor/ACC Board of Governors

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: 12, 2012

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? 04, 2018

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 (cardiosource.org) to include the macro-specifications 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.

Ad.8 Additional Information/Comments: Please note that the next scheduled review/update for this measure will occur at the same time as the new version release date of the registry in 2018.

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: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0642

Measure Title: Cardiac Rehabilitation Patient Referral From an Inpatient Setting

Measure Steward: American College of Cardiology

Brief Description of Measure: Percentage of patients admitted to a hospital with a primary diagnosis of an acute myocardial infarction or chronic stable angina or who during hospitalization have undergone coronary artery bypass (CABG) surgery, a percutaneous coronary intervention (PCI), cardiac valve surgery (CVS), or cardiac transplantation who are referred to an early outpatient cardiac rehabilitation/secondary prevention program.

Developer Rationale: 1. Cardiac rehabilitation/secondary prevention programs (CR/SP) improve patient outcomes, including quality of life, function, recurrent myocardial infarction, and mortality.

2. CR/SP is underutilized with geographic variability and decreased participation by patients with economic disadvantages, women and older patients.

3. The CR/SP performance measures were developed for use in systematic quality improvement projects to close this treatment gap.

4. Use of systematic referral processes and tools have been shown to increase CR/SP referral.

5. Enrollment and participation in CR/SP, not referral, have been shown to improve patient outcomes. However, referral is necessary for patients to enroll and participate in CR/SP. The strength of provider referral to CR has been shown to correlate with participation in CR.

6. Therefore, the specific CR/SP referral measures being submitted should be endorsed by NQF for use for quality improvement and in publicly reported systems.

Numerator Statement: Number of eligible patients with a qualifying event/diagnosis who have been referred to an outpatient Cardiac Rehabilitation/Secondary Prevention (CR/SP) program prior to hospital discharge or have a documented medical or system reason why such a referral was not made.

(Note: The program may include a traditional CR/SP program based on face-to-face interactions and training sessions or may include other options such as home-based approaches. If alternative CR/SP approaches are used, they should be designed to meet appropriate safety standards and deliver effective, evidence-based services.)

Denominator Statement: Number of hospitalized patients in the reporting period hospitalized with a qualifying cardiovascular disease event/diagnosis who do not meet any of the criteria listed in the denominator exclusion section below.

Denominator Exclusions: Exceptions criteria require documentation of one or more of the following factors that may prohibit cardiac rehabilitation participation:

-Medical factors (e.g., patient deemed by provider to have a medically unstable, life-threatening condition).

-Health care system factors (e.g., no cardiac rehabilitation/secondary prevention (CR/SP) program available within 60 min of travel time from the patient's home).

The only exclusion criterion for this measure is noted below:

-Patients who expired before discharge.

Measure Type: Process

Data Source: Electronic Health Records, Registry Data, Paper Medical Records

Level of Analysis: Clinician : Individual, Clinician: Group/Practice, Facility

IF Endorsement Maintenance – Original Endorsement Date: May 05, 2010 **Most Recent Endorsement Date:** Sep 08, 2014

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Summary of prior review in 2014

- The developer provides a diagram of referral to cardiac rehabilitation program and its relation to patient's health outcomes: Lower Mortality/Morbidity, Higher Quality of Life, Risk Factor Modification, Improved Function & Exercise Capacity, Improved Medication Adherence, Reduction in Re-Hospitalization Rates, and Cost Effective Care

- The developer cited systematic reviews of six ACCF/AHA guidelines with grading of the evidence for referral to cardiac rehabilitation for different heart disease/conditions. No QQC is provided for each of the six guidelines, but evidence grades are defined.
- The developer cited a Cochrane systematic reviews from 2009 with QQC provided and no evidence of publication bias for total mortality, CV mortality, CABG or PTCA. There was evidence of small study bias for total MI. Benefits are reported as risk ratios (95% CI) that compared participation in CR versus usual care based on meta-analyses from the Cochrane Systematic Review. Those with p value < 0.05 are:
    - For CR versus Usual Care, Follow-up > 12 months- Total Mortality (risk ratio=0.87) and CV Mortality (risk ratio=0.74)
    - For Hospital Admissions versus Usual Care, Follow-up 6-12 months with risk ratio of 0.69

Changes to evidence from last review

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

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

Updates:

- Although the developer attested there are no changes in evidence since it was endorsed in 2014, a new systematic review guideline and a new study from the Cochrane systematic review was added to the submission listed below:
- An AHA/ACC guideline for the management of patients with non-ST-elevation (NSTE) acute coronary syndromes (ACS). This guideline recommends all eligible patients with NSTE-ACS should be referred to a comprehensive cardiovascular rehabilitation program either before hospital discharge or during the first outpatient visit.
- A new study conducted from the Cochrane systematic review supports the conclusions of the prior review in 2014 that, compared with no exercise control, exercise-based CR reduces the risk of cardiovascular mortality but not total mortality.

Exception to evidence: N/A

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

Questions for the Committee:

- *The evidence provided by the developer is updated and directionally the same compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance from the Evidence Algorithm

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

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provides 2015-2016 data from two registries by quarterly and by decile.

1. ACTION Registry quarterly shown below:

Year 2015	Number of Hospitals	Percentage	Mean
Quarter 1	1028	79.05%	0.6117
Quarter 2	1048	78.877%	0.614845
Quarter 3	1065	78.11%	0.6166
Quarter 4	1100	77.79%	0.619972

Year 2016	Number of Hospitals	Percentage	Mean
Quarter 1	1101	77.5%	0.613781
Quarter 2	1111	77.72%	0.6185
Quarter 3	1119	78.19%	0.62209
Quarter 4	1122	78.55%	0.626772

2. CathPCI Registry quarterly shown below:

Year 2015	Number of Hospitals	Percentage	Mean
Quarter 1	1746	62.49%	0.529481
Quarter 2	1759	61.82%	0.5315
Quarter 3	1755	61.32%	0.529663
Quarter 4	1775	61%	0.527809

Year 2016	Number of Hospitals	Percentage	Mean
Quarter 1	1794	61.25%	0.53779
Quarter 2	1798	61.48%	0.535781
Quarter 3	1725	61.73%	0.537336
Quarter 4	1741	61.78%	0.538281

Disparities

- The developer provides 2012 data from two registries (ACTION and CathPCI) by gender, race, insurance, hospital teaching status, and hospital community.
- The data showed that Cardiac rehabilitation/secondary prevention programs (CR/SP) improve patient outcomes, including quality of life, function, recurrent myocardial infarction, and mortality. However, CR/SP is underutilized with geographic variability and decreased participation by patients with economic disadvantages, specifically women and older patients. The CR/SP performance measures were developed for use in systematic quality improvement projects to close this treatment gap.
- Use of systematic referral processes and tools have been shown to increase CR/SP referral. Enrollment and participation in CR/SP, not referral, have been shown to improve patient outcomes. However, referral is

necessary for patients to enroll and participate in CR/SP. The strength of provider referral to CR has been shown to correlate with participation in CR.

Questions for the Committee:

- *Does the data demonstrate a quality problem related to patients that do not receive a referral to cardiac rehabilitation from an inpatient setting after a cardiac event?*
- *Is a national performance measure still warranted?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

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

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

1a. Evidence:

- This measure was initially endorsed in 2010 and most recently in 2014. The developer states there is no additional evidence but in fact there was an additional supportive Cochrane review in 2014.
- Originally endorsed in 2010 and supported again in 2014. This is a process measure deserving continued endorsement.
- New evidence supports existing evidence.
- This measure was originally endorsed in 2010 and 2014. The evidence has not changed substantially since then despite a new Cochrane Review. Evidence is moderate.
- This is a process measure that does not directly relate to the outcomes of patients hospitalized for AMI or angina who have undergone a significant cardiac revascularization, valve repair or transplantation. This suggests a more heterogeneous population of patients whose physiology is likely to be different. This measure has been in use for the past seven years. However at a population level the greater the engagement in cardiac rehab the better the outcomes. So it is very much dependent upon a sufficient volume to mimic the population as a whole that makes this measure appropriate.
- This is a maintenance process measure solely based on referral. The evidence does support the direct relationship between the measure and the benefit.
- This is a maintenance measure. There is a new/updated systematic review guideline included that addresses NSTE ACS patients. There is also a new Cochrane systematic review. Both are in the same direction as the original evidence (benefits of CR) and further support the measure. Measure numerator does not speak to Acute Coronary Syndrome that does not result in MI and received medical management
- Evidence is not strong...would call this moderate evidence
- Empirical evidence is strong for the process (Cardiac Rehab/Secondary Prevention (CR/SP)) measure supported by primary studies, systematic reviews, multiple guidelines, and a Cochrane review. The evidence applies directly to the measure. The process measure of referring these CV patients (AMI, chronic stable angina, CABG PCI, CVS, heart transplant) to CR/SP programs is directly related to related to patients enrolling and participating in CR/SP programs. While referral has not been shown to improve patient outcomes, it is necessary in order to garner enrollment and participation of course into CR/SP programs. It does appear the evidence supports CV mortality reductions, readmission reductions, possible benefits (trends) in total mortality reductions and other benefits (QOL) of enrollment into CR/SP programs.

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

- While referral to CR rehab appears to be steady in the high70% range, there is potential for improvement.
- A 2014 Cochrane review provides addition support for this measure. There is much room for improvement of this measure. Which is publicly reported
- Gaps exist in both registries that were tested. Race, sex, hospital type, and insurance disparities exist.
- A performance gap exists. Opportunity for improvement is moderate.

- Studies have demonstrated that there is a low referral rate to ambulatory cardiac rehab. So it is an assumption that general increased rates will improve all populations: JACC: Cardiovascular Interventions Volume 9, Issue 5, 14 March 2016, Pages 496-498
- The performance gap for this metric justifies such a performance measure. Disparities were addressed but beyond economic disadvantage is the issue of reliable transportation. Some regions may have CR facilities available that do not fall outside the exclusion/exception range but where a practical barrier still exist for patients and so they may decline or influence referral patterns for CR.
- Recent performance data is provided. In both registries presented, the progress toward improved adherence seems static. What was the performance gap in the measure presentation in 2010? If there has been no improvement, there may be other issues besides non-compliance with guidelines (eg. availability of programs to refer to - yes there may be a program within 60 min commute, but is it full? Is 60 min reasonable for patients, many of whom will have driving restrictions for varying amounts of time). Disparities data is presented - women, older patients, and economically disadvantaged were referred less often.
- There is a moderate gap
- While rates of performance have been slowly but steadily increasing, there is still a large gap in performance, especially noting that several years of process and quality improvement have already likely been undertaken in these registry hospitals/systems that have been studied. There is nevertheless still a large gap for improvement in these hospitals and a larger gap in systems not yet focusing on this topic. Disparity data is presented and it does appear that several groups have noted disparities for care in this area. As Quality improves for the measure over time, it is likely that the disparity gaps would be closed or minimized. It would be good for the developer to provide to the committee the specific groups/subtypes of opportunities for improvement of why systems/clinicians are not meeting the measure. Is this because of poor documentation of the systems for patients meeting the exception criteria of time/distance or having a medical issue for which they cannot be enrolled? Or is it simply that the referrals are not being made in the 20-40% of patients with opportunity for improvement? in short, section 1b3 should/could also include the specific reasons for opportunities for improvement from these data and registries. Also it appears some studies support that provider support facilitating enrollment to CR was more impactful than the strength of the referral (Mitoff PR et al Rehabil Nurs. 2005 Jul-Aug;30(4):140-6. Patient-provider communication regarding referral to cardiac rehabilitation.) if the developer can comment

Criteria 2: Scientific Acceptability of Measure Properties

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: NQF Staff

Evaluation of Reliability and Validity

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Staff Evaluation of Scientific Acceptability

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.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 0642

Measure Title: Cardiac Rehabilitation Patient Referral From an Inpatient Setting

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
   NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.

   TIPS: Consider the following: 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?

   ☒Yes (go to Question #2)

   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

   TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

   ☒Yes (go to Question #4)

   ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

3. Was **empirical VALIDITY testing** of patient-level data conducted?

   ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

   ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)

4. Was reliability testing conducted with computed performance measure scores for each measured entity?

   TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

   ☒Yes (go to Question #5)

   ☐No (go to Question #8)

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

   TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

   ☒Yes (go to Question #6)

   ☐No (please explain below then go to Question #8)

6. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

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

   ☐High (go to Question #8)

   ☒Moderate (go to Question #8)

   ☐Low (please explain below then go to Question #7)

7. Was other reliability testing reported?

   ☒Yes (go to Question #8)

   ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

8. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

   TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

   ☒Yes (go to Question #9)

   ☐No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)

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

   TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

   Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

   ☒Yes (go to Question #10)

   ☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

10. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

    ☒Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

    ☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

    ☐Insufficient (go to Question #11)

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

1. Were all potential threats to validity that are relevant to the measure empirically assessed?

   TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

☒Yes (go to Question #2)

☐No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

2. Analysis of potential threats to validity: Any concerns with measure exclusions?

   TIPS: Consider the following: 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)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

   ☒Yes (please explain below then go to Question #3)

   ☐No (go to Question #3)

   ☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

   The developer highlighted that the documentation and assessment of exclusions help mitigate potential bias in reporting (i.e., excluding patients who are actually eligible for cardiac rehabilitation referral, in order to improve performance scores). Based on the results of the Cardiac Rehabilitation Referral Reliability Testing (CR3) project, the time and effort to assess exclusions does not appear to add significant burden.

3. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

   ☒Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

   a. Is a conceptual rationale for social risk factors included? ☐Yes ☐No

   b. Are social risk factors included in risk model? ☐Yes ☐No

   c. Any concerns regarding the risk-adjustment approach?

   *TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? 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, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

   ☐Yes (please explain below then go to Question #4)

   ☐No (go to Question #4)

4. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

   ☒Yes (please explain below then go to Question #5)

   ☐No (go to Question #5)

   As documented in the datasets, the developer noted there is wide variation in performance for this measure, specifically variation in the delivery of cardiac rehabilitation referral. This variation in data helped to identified noted gaps in care and areas where improvement in care are needed.

5. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

☐Yes (please explain below then go to Question #6)

☒No (go to Question #6)

☐Not applicable (go to Question #6)

6. Analysis of potential threats to validity: Any concerns regarding missing data?

   ☐Yes (please explain below then go to Question #7)

   ☒No (go to Question #7)

ASSESSMENT OF MEASURE TESTING

7. Was empirical validity testing conducted using the measure as specified and appropriate statistical test?

   Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

   ☐Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

   ☒No (please explain below then go to Question #8)

8. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

   TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

   ☒Yes (go to Question #9)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

9. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

   ☐Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☒ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

10. Was validity testing conducted with computed performance measure scores for each measured entity?

    TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

    ☐Yes (go to Question #11)

    ☐No (please explain below and go to Question #13)

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

    TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

    ☐Yes (go to Question #12)

    ☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

12. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #14)

    ☐Moderate (go to Question #14)

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

13. Was other validity testing reported?

    ☐Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

14. Was validity testing conducted with patient-level data elements?

    TIPS: Prior validity studies of the same data elements may be submitted

    ☐Yes (go to Question #15)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

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

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #16)

    ☐No (please explain below and rate Question #16 as INSUFFICIENT)

16. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

    ☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

    ☐Insufficient (go to Question #17)

17. OVERALL VALIDITY RATING

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

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the

score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

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

2a1. Reliability-Specifications:

- Moderate reliability.
- Moderate
- The reliability specs are clearly defined
- Reliability does not seem to be an issue as the measure is straight forward with minimal exclusions to create reliability issues.
- Due to the number of included diagnosis eligible for referral there is some concern that all of the measured denominator codes may be accurately attributed.
- Accept staff determination that measure meets the criteria
- Concerns with reliability....moderate
- No concerns for reliability specifications

2a2. Reliability testing:

- No concerns
- None
- IRR of data elements had a kappa of at least 0.7. Signal to noise ratios were quite high.
- Accept staffs' preliminary assessment. No concerns
- Concerns with testing...moderate
- No concerns on reliability testing

2b2. Validity testing & 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data):

- Validity remains at moderate level for this process measure
- Moderate
- Face validity was approximately 4 out of 5, for agreement/strong agreement.
- No threats to validity
- Previous reviews have judged the validity to be "moderate" which is a judgment that I would endorse
- There could be missing data not due to outcome referral, but to improper attribution based on wide variety of eligible conditions as above.

- Low validity
- 2b1 and 2b4-7. No concerns on validity or threats to validity. 2B4. The data support that if the referral is made and made in strong way, patients are likely to enroll in CR/SP programs and have improved outcomes because of this. 2b6. Minimal missing data as there is only one exclusion. It appears as if maintenance measures now requires empirical validity testing now which the developer aims to test in the future. Validity explanation satisfies my concerns here however. Face validity is good and predictive validity appears to be reasonable given the strong evidence for the measure.

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

- The measure is not risk adjusted.
- The greatest risk adjustment would be assigning a risk to any of categories of heart disease for which the measure was designed to review referral rates. Further there is no standardization for what is an appropriate cardiac rehab. So the outcomes would not be broken out by type of heart disease compared to what type of cardiac rehab program. This raises concerns for me in evaluating the results.
- There are demographic elements recorded which will be beneficial in risk adjustment but this may not be as exhaustive as necessary to account for variability due to distance from a CR facility.
- 2b2 - I wonder how the "Available CR program within 60 min travel" is applied. To me the data would be more consistent if there was a certain number of miles used rather than time. It's easier to draw a circle on a map indicating the number of miles to a program than to figure out the travel time.
- 2b2-3. No other clear apparent other threats to validity. Exclusions are straight forward and no risk adjustment is needed. 2b2. Exclusions are in line with evidence and common sense 2b3 no risk adjustment needed

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent
3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All the data elements are captured in electronic clinical data. However, the developer states that the data are abstracted from a record by a third party other than the individual obtaining the original information.
- Developers noted two challenges to the feasibility of the measures and provided the ongoing solutions below:
    1. For the ACC CathPCI registry, referral to CR/SP following percutaneous intervention is low in comparison to other performance measures, but can be improved with education and improved processes. AACVPR have been working with the Society for Cardiovascular Angiography and Intervention (SCAI) and ACC to provide educational materials for cardiologists and their patients about the benefits of CR/SP, but no recent updates are provided in regards to the educational materials.
    2. For the ACTION registry, inter-rater reliability testing was not strong compared to the CR3 project, which may be due to the additional education needed for the implementation of the CR/SP referral measures. The measure testing workgroup plans to work with AACVPR, ACC, and AHA leadership to develop implementation notes to instruct abstractors and providers about the documentation details needed to meet the 3 components of CR/SP referral criteria.
- Licensing and fees associated with this measure from ACCF's Nation Cardiovascular Data Registry (NCDR) 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?*

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

Committee Pre-evaluation Comments: Criteria 3: Feasibility

3. Feasibility:

- Feasibility is moderate
- Moderate
- There is apparently an issue with recording data elements.
- I have no concern with feasibility
- Generally, there is good feasibility for this measure except that referral is often one of the most inconsistently recorded measures. Registries and EHRs may not capture referrals not done electronically through the EHR or via phone where a discrete field has not been recorded.
- Registry data, documentation of referrals in the medical record. There may be multiple ways a referral is initiated. Are all captured?
- Moderate feasibility
- No concerns on feasibility. Although concerns are noted from the developer, this appears to be more of an educational opportunity on patients undergoing PCI from the ACC CathPCI registry and possible abstractor and provider education in the ACTION registry

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- This measure is in use for Professional Certification or Recognition Program ACCTION Registry Achievement Award. This measure is also in use by three quality improvement programs for benchmarking or specific to an organization. The quality improvement programs are: (1) NCDR CathPCI registry, (2) NCDR ACTION registry, and (3) ACC Patient Navigator.
- The developer indicate planned use is public reporting and hope to expand the use of this measure in other payment programs (e.g., accountable care organizations, Medicare Advantage insurance plans, other health plans on the insurance marketplace).
- The developer notes that ACC has made a decision to voluntarily public report out of the ACTION registry.

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

measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

1. Performance results are delivered to CathPCI and ACTION registry who received quarterly benchmark reports.
   - Users for the CathPCI registry reported that some sites have expressed difficulty with identifying certain data elements due to how data is being currently captured. For example, communication of patient details to CR facilities is sometimes challenging to capture. However there are many facilities that seem to have very good processes in place that are integrated with their EHR/EMR.
   - Users from the ACTION registry reported challenges related to sites being able to implement a process at their facility to streamline compliance to the measure since it requires a multi-provider approach to complete the process. However, many sites have been able to develop quality improvement initiatives to improve their compliance.
   - This measure have not been modified since last endorsement on September 8, 2014.

Additional Feedback: N/A

***Questions for the Committee**:*

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

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

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

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

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

Improvement results

- ACTION registry shows modest improvements from the performance data 2015-2016. The mean performance for Q4 of 2015 is 62% and the mean performance of Q4 in 2016 is 62.7%. The IQR for the ACTION registry shows a smaller range from 40.3% in Q4 of 2015 to 36.9% in Q4 of 2016.
- The CathPCI registry is consistent with the ACTION registry in showing modest improvements from 2015-2016 as well. The mean performance for quarter 4 of 2015 is 52.8% and the mean performance of quarter 4 of 2016 is 53.8%. The IQR for the CathPCI registry also shows a smaller range from 74.7% in quarter 4 of 2015 to 67% in quarter 4 of 2016.
- Developer believe the expanded use of this measure "will lead to greater awareness and accountability among providers and accelerate improvements in referral (and enrollment) rates".

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

Unexpected findings (positive or negative) during implementation

- No unexpected negative findings are reported.
- The developers noted unexpected benefits from two registries mentioned below:
  1. For CathPCI: An unexpected/good outcome of this measure is that many facilities reexamined their processes to be in compliance with CR parameters and found that their process needed improvement.
  2. For ACTION: An unexpected benefit is the improved patient compliance and commitment for other cardiac care measures which has a positive impact the long term outcomes of the patient.

Potential harms

- The developer did not report any unintended consequences.

Additional Feedback:

- Based on the NQF's Cardiovascular report in 2014, the Committee was very supportive of the importance of cardiac rehabilitation for this subset of patients and noted that multiple studies have shown reduction in both total and cardiac mortality in CHD patients after cardiac rehabilitation. The Committee did note the measure performance of the measure in CATH PCI Registry and ACTION Registry where there could be significant opportunity for improvement. The Committee recommended that the developer strengthen the measure by coupling referral with counseling the patient about the value of cardiac rehabilitation.

***Questions for the Committee**:*

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

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

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

4a. Use:

- This measure is publicly reported and part of an accountability program.
- It is useful measure
- This measure is in use for Professional Certification or Recognition Program ACCTION Registry Achievement Award, NCDR CathPCI registry, NCDR ACTION registry, and the ACC Patient Navigator. It is publicly reported by the ACTION registery.
- As above in 2b3, I am concerned with both lack of primary endpoint making this a process measure; then with the lack of consistency of inclusion in the numerator the type of heart disease and the characteristics of the rehab program
- Currently the metric is being used for the ACTION award certification and is publicly reported. Feedback has been provided informally. No formal feedback process identified.
- Has been used successfully.
- Not much data...very disturbing
- Current use of the measure is publicly reported by ACC (ACTION registry info) and also used in an accountability program and three quality improvement programs. Developer indicates that there are plans to expand the use to other payment programs such as ACO, Medicare advantage, and other health plans.

4b. Usability:

- As above.
- No apparent harm
- Used for quality improvement
- The measure is usable to drive referral to CR/SP. No harms have been identified.
- It may be difficult to gauge the conclusion in this heterogeneous population
- No unintended harms other than lower reimbursement potentially for facilities that do not adequately document referral to a CR program. Generally should lead to providing a streamlined referral process in a systematic manner for patients that could benefit.
- Moderate usability and use
- Agree with the developer that expanded use of the measure would lead to a greater awareness and accountability and accelerate improvements in referral and enrollment rates and subsequently improve care in

these patient groups. No concerns of harm from this measure. Appears benefits strongly outweigh harms throughout.

Criterion 5: Related and Competing Measures

Related or competing measures

- There may be related measures and are listed below:
    1. 0071: Persistence of Beta-Blocker Treatment After a Heart Attack
    2. 0090: Emergency Medicine: 12-Lead Electrocardiogram (ECG) Performed for Non-Traumatic Chest Pain
    3. 0137: ACEI or ARB for left ventricular systolic dysfunction- Acute Myocardial Infarction (AMI) Patients
    4. 0142: Aspirin prescribed at discharge for AMI
    5. 0230: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older.
    6. 0290: Median Time to Transfer to Another Facility for Acute Coronary Intervention
    7. 0643: Cardiac Rehabilitation Patient Referral From an Outpatient Setting (in our portfolio)
    8. 0730: Acute Myocardial Infarction (AMI) Mortality Rate
    9. 0964 Therapy with Aspirin, P2Y12 Inhibitor, and Statin at Discharge Following PCI in Eligible Patients
    10. 2377: Defect Free Care for AMI
    11. 2379: Adherence to Antiplatelet Therapy after Stent Implantation
    12. 2452 PCI: Post-Procedural Optimal Medical Therapy [clinician]
    13. 2473: Hospital 30-Day Risk-Standardized Acute Myocardial Infarction (AMI) Mortality eMeasure
- There are no competing measures.

Harmonization

1. Committee recommendations for combining or harmonizing measures may be solicited.

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

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 10, 2018

No comments have been submitted as of this date.

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

0642_NQF_evidence_attachment_Sep2017_v2.pdf

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission? Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1a Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0642

Measure Title: Cardiac Rehabilitation Patient Referral From an Inpatient Setting

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 11/8/2017

Instructions

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

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

1a. Evidence to Support the Measure Focus

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

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.

- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

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

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

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

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

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

Outcome

☐ Outcome:

☐Patient-reported outcome (PRO):

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

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

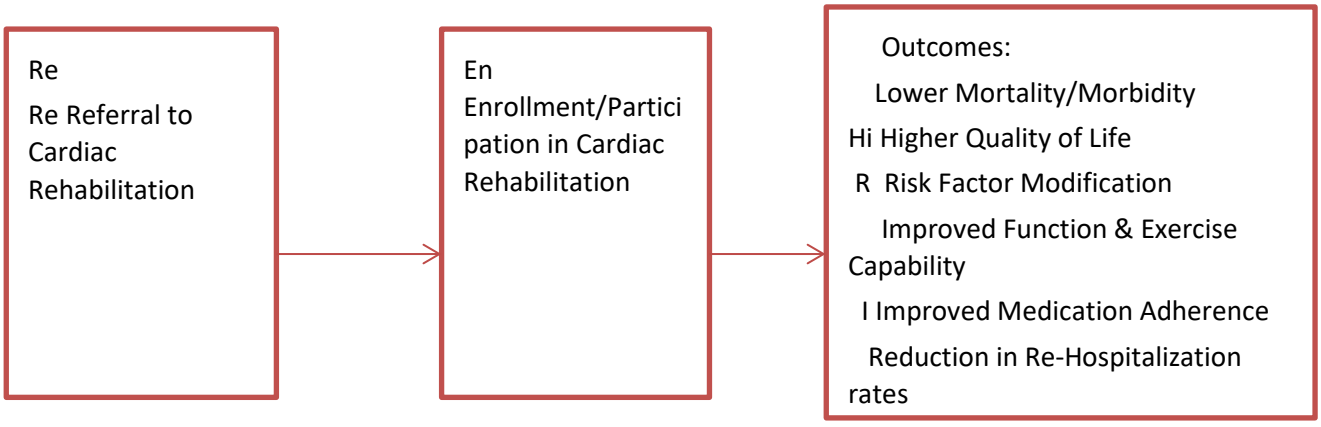
☒ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.



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

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Coronary Artery Bypass Surgery (CABG): Hillis LD, et. Al. 2011 ACCF/AHA guideline for coronary artery bypass graft surgery: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2011;124:e652– e735. http://circ.ahajournals.org/content/124/23/e652.full.pdf+html
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Pages e683-684: 4.9. Cardiac Rehabilitation: Cardiac rehabilitation is recommended for all eligible patients after CABG.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	Cardiac rehabilitation is recommended for all eligible patients after CABG. Recommendation Class I; Level of Evidence: A
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Percutaneous Coronary Intervention (PCI): Levine GN, 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. *J Am Coll Cardiol* 2011;58:e44–122. http://content.onlinejacc.org/article.aspx?articleid=1147816
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Page e89: 6.4.3. Cardiac Rehabilitation: Recommendation Medically supervised exercise programs (cardiac rehabilitation) should be recommended to patients after PCI, particularly for moderate- to high-risk patients for whom supervised exercise training is warranted.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	Medically supervised exercise programs (cardiac rehabilitation) should be recommended to patients after PCI, particularly for moderate- to high-risk patients for whom supervised exercise training is warranted. Recommendation Class I; Level of Evidence: A
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Coronary Artery Disease (CAD): Smith SC Jr., et. Al. 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
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Page 2436 1. All eligible patients with ACS or whose status is immediately post coronary artery bypass surgery or post-PCI should be referred to a comprehensive outpatient cardiovascular rehabilitation program either prior to hospital discharge or during the first follow-up office visit. 2. All eligible outpatients with the diagnosis of ACS, coronary artery bypass surgery or PCI, chronic angina… within the past year should be referred to a comprehensive outpatient cardiovascular rehabilitation program. 3. A home-based cardiac rehabilitation program can be substituted for a supervised, center-based program for low-risk patients.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	All eligible patients with ACS or whose status is immediately post coronary artery bypass surgery or post-PCI should be referred to a comprehensive outpatient cardiovascular rehabilitation program either prior to hospital discharge or during the first follow-up office visit. Recommendation Class I; Level of Evidence: A All eligible outpatients with the diagnosis of ACS, coronary artery bypass surgery or PCI, chronic angina, and/or peripheral artery disease within the past year should be referred to a comprehensive outpatient cardiovascular rehabilitation program. PCI: Recommendation Class I; Level of Evidence: A Chronic Angina: Recommendation Class I; Level of Evidence: B A home-based cardiac rehabilitation program can be substituted for a supervised, center-based program for low-risk patients. Recommendation Class I; Level of Evidence: A
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	ST Elevation Myocardial Infarction (STEMI): O'Gara PT, 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:e78 –140, doi:10.1016/j.jacc.2012.11.019. http://content.onlinejacc.org/article.aspx?articleid=1486115
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	e114-116: 11.1. Post hospitalization Plan of Care: Recommendations 1. Post hospital systems of care designed to prevent hospital readmissions should be used to facilitate the transition to effective, coordinated outpatient care for all patients with STEMI. 2. Exercise-based cardiac rehabilitation/secondary prevention programs are recommended for patients with STEMI. 3. A clear, detailed, and evidence-based plan of care that promotes medication adherence, timely follow- up with the healthcare team, appropriate dietary and physical activities, and compliance with interventions for secondary prevention should be provided to patients with STEMI
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	Post-hospital systems of care designed to prevent hospital readmissions should be used to facilitate the transition to effective, coordinated outpatient care for all patients with STEMI. Recommendation Class I; Level of Evidence: B Exercise-based cardiac rehabilitation/secondary prevention programs are recommended for patients with STEMI. Recommendation Class I; Level of Evidence: B A clear, detailed, and evidence-based plan of care that promotes medication adherence, timely follow-up with the healthcare team, appropriate dietary and physical activities, and compliance with interventions for secondary prevention should be provided to patients with STEMI. Recommendation Class I; Level of Evidence: C
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Amsterdam EA, Wenger NK, Brindis RG, 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. J Am Coll Cardiol. 2014;64:e139-228/ http://www.onlinejacc.org/content/64/24/e139.
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	e139-228 Cardiac Rehabilitation and Physical Activity: Recommendation: All eligible patients with NSTE-ACS should be referred to a comprehensive cardiovascular rehabilitation program either before hospital discharge or during the first outpatient visit. (Class I, Level of Evidence: B)
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	All eligible patients with NSTE-ACS should be referred to a comprehensive cardiovascular rehabilitation program either before hospital discharge or during the first outpatient visit. (Class I, Level of Evidence: B)
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Stable Ischemic Heart Disease (Stable IHD): Fihn SD, 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, 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://circ.ahajournals.org/content/early/2012/11/19/CIR.0b013e318277d6a0.full.pdf
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Pages e91-92: 4.4.1.4. PHYSICAL ACTIVITY 1. For all patients, the clinician should encourage 30 to 60 minutes of moderate-intensity aerobic activity, such as brisk walking, at least 5 days and preferably 7 days per week, supplemented by an increase in daily lifestyle activities (e.g., walking breaks at work, gardening, household work) to improve cardiorespiratory fitness and move patients out of the least-fit, least-active, high-risk cohort (bottom 20%). 3. Medically supervised programs (cardiac rehabilitation) and physician-directed, home-based programs are recommended for at risk patients at first diagnosis.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	For all patients, the clinician should encourage 30 to 60 minutes of moderate-intensity aerobic activity, such as brisk walking, at least 5 days and preferably 7 days per week, supplemented by an increase in daily lifestyle activities (e.g., walking breaks at work, gardening, household work) to improve cardiorespiratory fitness and move patients out of the least-fit, least-active, high-risk cohort (bottom 20%). Recommendation Class I; Level of Evidence: B Medically supervised programs (cardiac rehabilitation) and physician-directed, home-based programs are recommended for at risk patients at first diagnosis. Recommendation Class I; Level of Evidence: A
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Table 1 (applies to all Provide all other grades and definitions from the evidence grading system)

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 a 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 considered may/might be reasonable usefulness/effectiveness is unknown/unclear/ uncertain or not well established	COR III: No Benefit — COR III: Harm is not recommended — potentially harmful is not indicated — causes harm should not be performed/ administered/ other — associated with excess morbidity/ mortality is not useful/ beneficial/ effective — 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.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Heran BS,et al. Exercise-based cardiac rehabilitation for coronary heart disease. *Cochrane Database of Systematic Reviews* 2011, Issue 7. Art. No.: CD001800. DOI: 10.1002/14651858.CD001800.pub2. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001800.pub2/pdf **Citation and URL for methodology for evidence review and grading**: The systematic review identified quality of evidence based on risk of bias. System for determining risk of bias was explained in Chapter 8 of Cochrane Handbook for Systematic Reviews for Interventions, 5.0.2, updated September 2009 http://www.mrc-bsu.cam.ac.uk/cochrane/handbook502 .
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**The information in the following questions in this section is based on the Cochrane Systematic Review cited in the source of the systematic review.** **Intervention/Service**: The effectiveness of exercise-based cardiac rehabilitation on mortality, morbidity and health-related quality of life of patients with CHD is addressed. Exercise-based cardiac rehabilitation is defined as a supervised or unsupervised inpatient, outpatient, or community- or home-based intervention including some form of exercise training that is applied to a cardiac patient population. The intervention could be exercise training alone or exercise training in addition to psychosocial and/or educational interventions (i.e. “comprehensive cardiac rehabilitation”). Usual care could include standard medical care, such as drug therapy, but did not receive any form of structured exercise training or advice. **Outcomes**: Total mortality; Total MI; Total revascularizations; Total hospitalizations; Health-related quality of life; Costs and cost-effectiveness

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	An overall grade of methodological quality was not assigned. In the systematic review, individual study quality was graded on a scale for risk of bias. **Allocation:** Nearly all the trial publications simply reported that the trial was "randomized" but did not provide any details. A total of 8/47 (17%) studies reported details of appropriate generation of the random sequence and 7/47 (15%) studies reported appropriate concealment of allocation. **Blinding:** For exercise-based cardiac rehabilitation trials, it is not possible to blind patients and clinicians to the intervention. For the large majority of studies, insufficient information was provided to evaluate the blinding of assessors; only 4 of 47 (9%) reported that outcome assessors were blind to group allocation. **Incomplete outcome data:** Losses to follow-up and drop out were relatively high, ranging from 21% to 48% in 12 trials. Follow-up of 80% or more was achieved in 33/47 (70%) studies. Furthermore, reasons for loss to follow and dropout were often not reported. Two trials did not report information on losses to follow-up. Several trials have excluded significant numbers of patients post-randomization, and thus in an intention to treat analysis, these then have been regarded as dropouts. **Selective reporting:** A number of the included studies were not designed to assess treatment group differences in morbidity and mortality (as these were not the primary outcomes of these trials) and, therefore, may not have fully reported all clinical events that occurred during the follow-up period. All studies collecting validated health-related quality of life outcomes fully reported these outcomes. **Quality of the evidence:** We found no evidence of publication bias for total mortality, CV mortality, CABG or PTCA. There was evidence of small study bias for total MI.
Provide all other grades and definitions from the evidence grading system	Two reviewers (BSH, JMHC) independently assessed the risk of bias in included studies using the Cochrane Collaboration's recommended tool, which is a domain-based critical evaluation of the following domains: sequence generation; allocation concealment; blinding of outcome assessment; incomplete outcome data; and selective outcome reporting. Only author's recommendations were provided: In medium to longer term (i.e. 12 or more months follow-up) exercise-based cardiac rehabilitation is effective in reducing overall and cardiovascular mortality and appears to reduce the risk of hospital admissions in the shorter-term (< 12 months follow-up) in patients with CHD. The available evidence does not demonstrate a reduction in the risk of total MI, CABG or PTCA with exercise based cardiac rehabilitation as compared to usual care at any duration of follow-up. Exercise-based cardiac rehabilitation should be recommended for patients similar to those included in the randomized controlled trials
Grade assigned to the **recommendation** with definition of the grade	NA
Provide all other grades and definitions from the recommendation grading system	NA

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity-Seventeen studies (26 publications) met the inclusion criteria and had extractable data to assess the effects of exercise-based cardiac rehabilitation, compared with usual care, on mortality and morbidity in patients with CHD. These were added to the 30 studies (55 publications) from the original Cochrane review for a total of 47 studies (81 publications). Randomized controlled trials (RCTs) of exercise-based cardiac rehabilitation versus usual care with a follow-up period of at least six months. A total of 47 RCTs, with 10,794 patients. Quality-Trial sample sizes varied widely from 28 to 2304, with a median intervention duration of three (range 0.25 to 30) months and a follow-up of 24 (range six to 120) months. Nearly all the trial publications simply reported that the trial was "randomized" but did not provide any details. For exercise-based cardiac rehabilitation trials, it is not possible to blind patients and clinicians to the intervention. For the large majority of studies, insufficient information was provided to evaluate the blinding of assessors; only 4 of 47 (9%) reported that outcome assessors were blind to group allocation. Losses to follow-up and drop out were relatively high, ranging from 21% to 48% in 12 trials. Follow-up of 80% or more was achieved in 33/47 (70%) studies. Based on funnel plot analysis, no publication bias was found for all cause mortality, cardiovascular mortality, CABG and PTCA. However, there appears to an absence of negative-result trials of small to medium size for MI which was statistically significant (P = 0.019).

Estimates of benefit and consistency across studies	Predictors of all-cause mortality, cardiovascular mortality, recurrent MI, and revascularisation (CABG and PTCA) were examined using univariate meta-regression. …a reduction in both total and cardiac mortality was observed in CHD patients randomized to exercise-based rehabilitation. However, this updated review shows that this mortality benefit is limited to studies with a follow-up of greater than 12months. We also found that with exercise the rate of hospital readmissions may be reduced in studies up to 12 months follow-up (based on 4 trials with 54/254 versus 73/225 events), but not in longer term follow-up. There was no difference between exercise-based cardiac rehabilitation and usual care groups in the risk of recurrent myocardial infarction or revascularization at any duration of follow-up. The following are risk ratios (95% CI); (p) comparing participation in CR versus usual care based on meta-analyses from the Cochrane Systematic Review. **Total Mortality**; CR vs Usual Care, Follow-up 6-12 months: 0.82 [0.67, 1.01]; (p = 0.061) **Total Mortality**; CR vs Usual Care, Follow-up >12 months: 0.87 [0.75, 0.99]; (p = 0.041); **CV Mortality**; CR vs Usual Care, Follow-up 6-12 months: 0.93 [0.71, 1.21]; (p=0.590) **CV Mortality**; CR vs Usual Care, Follow-up >12 months: 0.74 [0.63, 0.87]; (p= 0.00018) **Fatal and/or nonfatal MI** vs Usual Care, Follow –up 6-12 months: 0.92 [0.70, 1.22]; (p=0.560) **Fatal and/or nonfatal MI** vs Usual Care, Follow – up >12 months: 0.97 [0.82, 1.15]; (p=0.730) **CABG vs Usual Care**, Follow –up 6-12 months: 0.91 [0.67, 1.24]; (p=0.550) **CABG vs Usual Care**, Follow – up >12 months: 0.93 [0.68, 1.27]; (p=0.650) NQF staff enter #/title Version 6.5 05/29/13 9 **PTCA** vs Usual Care, Follow –up 6-12 months: 1.02 [0.69, 1.50]; (p=0.930) **PTCA** vs Usual Care, Follow – up >12 months: 0.89 [0.66, 1.19]; (p=0.420) **Hospital Admissions** vs Usual Care, Follow –up 6-12 months: 0.69 [0.51, 0.93]; (p=0.016) **Hospital Admissions** vs Usual Care, Follow – up >12 months: 0.98 [0.87, 1.11]; (p=0.790) Given both the heterogeneity in outcome measures and methods of reporting findings, a meta-analysis was not undertaken for **health-related quality of life**. In seven out of 10 trials reporting health related quality of life using validated measures there was evidence of a significantly higher level of quality of life with exercise-based cardiac rehabilitation than usual care

What harms were identified?	Although this review did not assess harm, "several studies have documented the safety of exercise based cardiac rehabilitation in patients with documented SIHD. The 2007 AHA Scientific Statement on Exercise and Acute Cardiovascular Events estimates the risk of a major adverse cardiac event (MACE) at 1 in 80,000 patient-hours. This low event rate applies to medically supervised programs that evaluate patients before participation, provide serial surveillance, and are equipped to handle emergencies." Fihn SD, 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, 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://circ.ahajournals.org/content/early/2012/11/19/CIR.0b013e318277d6a0.full.pdf
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Anderson L, Thompson DR, Oldridge N, Zwisler AD, Rees K, Martin N, Taylor RS. Exercise-based cardiac rehabilitation for coronary heart disease. *Cochrane Database of Systematic Reviews* 2016, Issue 1. Art. No.: CD001800. DOI: 10.1002/14651858.CD001800.pub3. From the study: This updated Cochrane review supports the conclusions of the previous version of this review that, compared with no exercise control, exercise-based CR reduces the risk of cardiovascular mortality but not total mortality.

1a.4 OTHER SOURCE OF EVIDENCE

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

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

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

1. Cardiac rehabilitation/secondary prevention programs (CR/SP) improve patient outcomes, including quality of life, function, recurrent myocardial infarction, and mortality.
2. CR/SP is underutilized with geographic variability and decreased participation by patients with economic disadvantages, women and older patients.
3. The CR/SP performance measures were developed for use in systematic quality improvement projects to close this treatment gap.
4. Use of systematic referral processes and tools have been shown to increase CR/SP referral.
5. Enrollment and participation in CR/SP, not referral, have been shown to improve patient outcomes. However, referral is necessary for patients to enroll and participate in CR/SP. The strength of provider referral to CR has been shown to correlate with participation in CR.
6. Therefore, the specific CR/SP referral measures being submitted should be endorsed by NQF for use for quality improvement and in publicly reported systems.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

2015-2016 Performance Rates (ACTION Registry) for Cardiac Rehab Inpatient Referral

Year 2016	Number of Hospitals	Numerator	Denominator	Percentage	Min	Mean	Max	IQR	Standard-Deviation
Q1	1101	119220	153834	0.775	0	0.613781	1	0.4051	0.409366757
Q2	1111	122828	158046	0.7772	0	0.6185	1	0.3763	0.408066401
Q3	1119	126384	161642	0.7819	0	0.62209	1	0.3617	0.403743484
Q4	1122	129231	164520	0.7855	0	0.626772	1	0.3694	0.401194211
Year 2015	**Number of Hospitals**	**Numerator**	**Denominator**	**Percentage**	**Min**	**Mean**	**Max**	**IQR**	**Standard-Deviation**
Q1	1028	113452	143516	0.7905	0	0.6117	1	0.4273	0.411990318
Q2	1048	114319	145121	0.7877	0	0.614845	1	0.415	0.409592333
Q3	1065	114989	147208	0.7811	0	0.6166	1	0.3873	0.410184958
Q4	1100	116774	150124	0.7779	0	0.619972	1	0.4025	0.406531493

2015-2016 Performance Rates (ACTION Registry) for Cardiac Rehab Inpatient Referral by Decile

Year 2016	0	5	10	15	25	50	75	85	90	95	100
Q1	0	0.0181	0.115	0.2863	0.5504	0.8482	0.9555	0.9828	0.9953	1	1
Q2	0	0.0047	0.1225	0.3138	0.5801	0.8497	0.9564	0.9819	0.9944	1	1
Q3	0	0.0229	0.1383	0.3112	0.5909	0.8503	0.9526	0.9821	0.9947	1	1
Q4	0	0.0258	0.1423	0.3485	0.5862	0.8581	0.9556	0.9834	0.9946	1	1
Year 2015	0	5	10	15	25	50	75	85	90	95	100
Q1	0	0.0201	0.1165	0.263	0.5316	0.8572	0.9589	0.986	0.9954	1	1
Q2	0	0.0237	0.1199	0.2807	0.5412	0.8606	0.9562	0.9855	0.9955	1	1
Q3	0	0.0147	0.1025	0.3056	0.5693	0.8531	0.9566	0.9851	0.9957	1	1
Q4	0	0.0186	0.1253	0.3212	0.5611	0.8471	0.9636	0.9853	0.9975	1	1

2015-2016 Performance Rates (CathPCI Registry) for Cardiac Rehab Inpatient Referral

Year 2016	Number of Hospitals	Numerator	Denominator	Percentage	Min	Mean	Max	IQR	Standard-Deviation
Q1	1794	391824	639751	0.6125	0	0.53779	1	0.6889	0.466717104
Q2	1798	396854	645452	0.6148	0	0.535781	1	0.7091	0.467904819
Q3	1725	400702	649104	0.6173	0	0.537336	1	0.6848	0.46616173
Q4	1741	405801	656858	0.6178	0	0.538281	1	0.6705	0.463936444
Year 2015	Number of Hospitals	Numerator	Denominator	Percentage	Min	Mean	Max	IQR	Standard-Deviation
Q1	1746	383291	613318	0.6249	0	0.529481	1	0.7609	0.475142364
Q2	1759	383985	621155	0.6182	0	0.5315	1	0.7558	0.475110579
Q3	1755	384829	627597	0.6132	0	0.529663	1	0.7548	0.474199117
Q4	1775	387736	635651	0.61	0	0.527809	1	0.7471	0.472158047

2015-2016 Performance Rates (CathPCI Registry) for Cardiac Rehab Inpatient Referral by Decile

Year 2016	0	5	10	15	25	50	75	85	90	95	100
Q1	0	0.0043	0.0168	0.0412	0.2486	0.7148	0.9375	0.9721	0.9856	0.9948	1
Q2	0	0.0032	0.0169	0.0414	0.2273	0.7197	0.9364	0.971	0.9836	0.9941	1
Q3	0	0.0041	0.015	0.0418	0.2488	0.7216	0.9336	0.9689	0.9829	0.994	1
Q4	0	0.0044	0.0164	0.0474	0.26	0.722	0.9305	0.9664	0.9809	0.9931	1
Year 2015	**0**	**5**	**10**	**15**	**25**	**50**	**75**	**85**	**90**	**95**	**100**
Q1	0	0.0008	0.0074	0.0264	0.1787	0.7131	0.9396	0.9749	0.9867	0.9967	1
Q2	0	0.0006	0.0086	0.0254	0.1882	0.7221	0.944	0.9752	0.9866	0.9958	1
Q3	0	0.0008	0.0088	0.0266	0.1841	0.7114	0.9389	0.9728	0.987	0.9959	1
Q4	0	0.001	0.0104	0.0303	0.1838	0.6961	0.9309	0.972	0.9861	0.9953	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.

1. Thomas RJ, Miller NH, Lamendola C, Berra K, Hedbäck B, Durstine JL, Haskell W. National Survey on Gender Differences in Cardiac Rehabilitation Programs. Patient characteristics and enrollment patterns. J Cardiopulm Rehabil. 1996 Nov-Dec;16(6):402-12.

2. Centers for Disease Control and Prevention (CDC). Receipt of outpatient cardiac rehabilitation among heart attack survivors--United States, 2005. MMWR Morb Mortal Wkly Rep. 2008 Feb 1;57(4):89-94.

3. Suaya J, Shepard DS, Normand SL, Ades PA, Prottas J, Stason WB. Use of cardiac rehabilitation by Medicare beneficiaries after myocardial infarction or coronary bypass surgery. Circulation. 2007 Oct 9;116(15):1653-62.

4. Curnier DY, Savage PD, Ades PA. Geographic distribution of cardiac rehabilitation programs in the United States. J Cardiopulm Rehabil. 2005 Mar-Apr;25(2):80-4.

5. Grace SL, Gravely-Witte S, Brual J, Monette G, Suskin N, Higginson L, Alter DA, Stewart DE. Contribution of patient and physician factors to cardiac rehabilitation enrollment: a prospective multilevel study. Eur J Cardiovasc Prev Rehabil. 2008 Oct;15(5):548-56

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

Geographic area: The CathPCI and ACTION Registry-GWTG collect 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: CathPCI: 1284 for calendar year 2011; 1360 for calendar year 2012:

Number of accountable entities: ACTION Registry-GWTG: 551 for calendar year 2011; 703 for calendar year 2012

Patients included: CathPCI: 223037 for calendar years 2011-2012;

Patients included: CathPCI: 1,239,643 for calendar years 2011-2012

Disparities by Gender 2012 (ACTIONRegistry-GWTG)

	Total	male	P-Value
	n = 122285	Male	
n = 81201		Female	
n = 41084			
CR			
Cardiac Rehab Referral	92362 (75.5%)	62725 (77.2%)	29637 (72.1%) < 0.001

Continuous variables compared using Student´s T-test.

Categorical variables compared using chi-square or Fisher´s exact test.

Disparities by Race 2012 (ACTIONRegistry-GWTG)

	Total	racecat	P-Value
	n = 122285	1 Caucasian	
n = 103641		2 Af Am	
n = 14329		3 Other	
n = 4315			
CR			
Cardiac Rehab Referral	92362 (75.5%)	79246 (76.5%)	10308 (71.9%) 2808 (65.1%) < 0.001

Continuous variables compared using one-way analysis of variance.

Categorical variables compared using chi-square or Fisher´s exact test.

Disparities by Insurance 2012 (ACTIONRegistry-GWTG)

	Total	inscat	P-Value
	n = 122285	1 Private	
n = 70170		2 Medicare	
n = 28803		3 Medicaid	
n = 5273		4 Other	
n = 2949		5 None	
n = 15090			
CR			
Cardiac Rehab Referral	92362 (75.5%)	54457 (77.6%)	20205 (70.1%) 3713 (70.4%) 2192 (74.3%) 11795 (78.2%) < 0.001

Continuous variables compared using one-way analysis of variance.

Categorical variables compared using chi-square or Fisher´s exact test.

Disparities by Hospital Teaching status 2012 (ACTIONRegistry-GWTG)

	Total	IsTeaching	P-Value
	n = 122285	Teaching Hosp	
n = 56023		Non-Teaching Hosp	
n = 66262			
CR			

Cardiac Rehab Referral 92362 (75.5%) 44626 (79.7%) 47736 (72.0%) < 0.001

Continuous variables compared using Student´s T-test.

Categorical variables compared using chi-square or Fisher´s exact test.

Disparities by Hospital Community2012(ACTIONRegistry-GWTG)

Total CommunityDesc P-Value

n = 122285 Rural

n = 17667 Suburban

n = 36800 Urban

n = 67818

CR

Cardiac Rehab Referral 92362 (75.5%) 13524 (76.5%) 27467 (74.6%) 51371 (75.7%) < 0.001

Continuous variables compared using one-way analysis of variance.

Categorical variables compared using chi-square or Fisher´s exact test.

Disparities by Gender 2012 (CathPCI)

Total Sex P-Value

n = 623098 Male

n = 424459 Female

n = 198639

C Rehab

Cardiac Rehabilitation Referral 383112 (61.49%) 261946 (61.71%) 121166 (61.00%) < 0.001

Continuous variables compared using Student´s T-test.

Categorical variables compared using chi-square or Fisher´s exact test.

Disparities by Race 2012 (CathPCI)

Total racecat P-Value

n = 623098 1 Caucasian

n = 542871 2 Af Am

n = 52261 3 Other

n = 27966

C Rehab

Cardiac Rehabilitation Referral 383112 (61.49%) 340224 (62.67%) 29994 (57.39%) 12894 (46.11%) < 0.001

Continuous variables compared using one-way analysis of variance.

Categorical variables compared using chi-square or Fisher´s exact test.

Disparities by Insurance 2012 (CathPCI)

Total inscat P-Value

n = 623098 1 Private

n = 399887 2 Medicare

n = 140623 3 Medicaid

n = 23515 4 Other

n = 14177 5 None

n = 44896

C Rehab

Cardiac Rehabilitation Referral 383112 (61.49%) 249706 (62.44%) 82008 (58.32%) 13741 (58.44%) 8689 (61.29%) 28968 (64.52%) < 0.001

Continuous variables compared using one-way analysis of variance.

Categorical variables compared using chi-square or Fisher´s exact test.

Disparities by Hospital Teaching status 2012 (CathPCI)

Total Teaching Hospital P-Value

n = 623098 1

n = 310334 0

n = 312764

C Rehab

Cardiac Rehabilitation Referral 383112 (61.49%) 191840 (61.82%) 191272 (61.16%) < 0.001

Continuous variables compared using Student´s T-test.

Categorical variables compared using chi-square or Fisher´s exact test.

Disparities by Hospital Community 2012 (CathPCI)

Total Hospital Location P-Value

n = 623098 RURAL

n = 81090 SUBURBAN

n = 190630 URBAN

n = 351378

C Rehab

Cardiac Rehabilitation Referral 383112 (61.49%) 51938 (64.05%) 118013 (61.91%) 213161 (60.66%) < 0.001

Continuous variables compared using one-way analysis of variance.

Categorical variables compared using chi-square or Fisher´s exact test.

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

1. Thomas RJ, Miller NH, Lamendola C, Berra K, Hedbäck B, Durstine JL, Haskell W. National Survey on Gender Differences in Cardiac Rehabilitation Programs. Patient characteristics and enrollment patterns. J Cardiopulm Rehabil. 1996 Nov-Dec;16(6):402-12.

2. Suaya Ja, Shepard Ds, Normand ST, Ades PA, Prottas J, Stason WB. Use of cardiac rehabilitation by Medicare beneficiaries after myocardial infarction or coronary bypass surgery. Circulation 2007;116:1653-1662.

3. Weingarten MN, Salz KA, Thomas RJ, Squires RW. Rates of enrollment for Men and Women Referred to Outpatient Cardiac Rehabilitation. J Cardiopulm Rehabil Prev. 2011 July/August;31(4):217-22.

4. Review article: Valencia HE, Savage PD, Ades PA. Cardiac rehabilitation participation in underserved populations. J Cardiopulm Rehabil Prev. 2011;31:203-210.

5. Beswick AD, Rees K, Griebsch I, Taylor FC, Burke M, West RR, Victory J, Brown J, Taylor RS, Ebrahim S. Provision, uptake and cost of cardiac rehabilitation programmes: improving services to under-represented groups. Health Technol Assess. 2004 Oct;8(41):iii-iv, ix-x, 1-152.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Cardiovascular, Cardiovascular : Coronary Artery Disease, Cardiovascular : Coronary Artery Disease (AMI), Cardiovascular : Coronary Artery Disease (PCI), Surgery : Cardiac Surgery

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

Primary Prevention

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

Elderly, Populations at Risk, Populations at Risk : Individuals with multiple chronic conditions

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://content.onlinejacc.org/article.aspx?articleid=1138518

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:** action_v2_codersdictionary_2-4-2--rebranded-__AND_cathpci_v4_codersdictionary_4-4.pdf

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

There have been no changes since our submission in 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) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Number of eligible patients with a qualifying event/diagnosis who have been referred to an outpatient Cardiac Rehabilitation/Secondary Prevention (CR/SP) program prior to hospital discharge or have a documented medical or system reason why such a referral was not made.

(Note: The program may include a traditional CR/SP program based on face-to-face interactions and training sessions or may include other options such as home-based approaches. If alternative CR/SP approaches are used, they should be designed to meet appropriate safety standards and deliver effective, evidence-based services.)

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

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

Qualifying events include all patients hospitalized with primary diagnosis of myocardial infarction (MI), chronic stable angina, or who during hospitalization have undergone coronary artery bypass graft surgery (CABG), percutaneous coronary intervention (PCI), cardiac valve surgery, and/or heart transplantation.

A referral is defined as an official communication between the healthcare provider and the patient to recommend and carry out a referral order to an early outpatient cardiac rehabilitation program. This includes the provision of all necessary information to the patient that will allow the patient to enroll in an early outpatient cardiac rehabilitation program. This also includes a communication between the healthcare provider or healthcare system and the cardiac rehabilitation program that includes the patient´s enrollment information for the program. A hospital discharge summary or office note may be potentially formatted to include the necessary patient information to communicate to the cardiac rehabilitation program [the patient´s cardiovascular history, testing, and treatments, for instance.] All communications must maintain appropriate confidentiality as outlined by the 1996 Health Insurance Portability and Accountability Act (HIPAA).

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

Number of hospitalized patients in the reporting period hospitalized with a qualifying cardiovascular disease event/diagnosis who do not meet any of the criteria listed in the denominator exclusion section below.

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

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

Patients with a qualifying event who are to be discharged for a short-term stay in an inpatient medical rehabilitation facility are still expected to be referred to an outpatient cardiac rehabilitation program by the inpatient team during the index hospitalization. This referral should be reinforced by the care team at the medical rehabilitation facility.

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

Exceptions criteria require documentation of one or more of the following factors that may prohibit cardiac rehabilitation participation:

-Medical factors (e.g., patient deemed by provider to have a medically unstable, life-threatening condition).

-Health care system factors (e.g., no cardiac rehabilitation/secondary prevention (CR/SP) program available within 60 min of travel time from the patient's home).

The only exclusion criterion for this measure is noted below:

-Patients who expired before discharge.

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

Exclusion:

There is only one exclusion criteria (patients who expired before discharge). This information is readily available within the medical record.

Exceptions:

All eligible patients who can participate in even a low intensity exercise program and who have the cognitive ability to carry out the individualized education and counseling to life-long secondary prevention efforts should be referred to cardiac rehabilitation/secondary prevention programs, because morbidity and mortality benefits extend to nearly all patient populations, regardless of age or co-morbidities. As a result, the exception examples included in the performance measure relate to either the patient's inability to attend an exercise program (due to physical or practical obstacles) or to cognitive deficits which make them unable to actively participate in exercise or to apply secondary prevention recommendations.

Examples, justification, and data collection issues for exceptions for this measure;

1. Medical factors (e.g., patient deemed by provider to have a medically unstable, life-threatening condition): Medically unstable, life-threatening conditions are contraindications to aerobic exercise and require medical efforts to stabilize and reverse those conditions, rather than efforts directed at secondary prevention of cardiovascular disease. Objective criteria for contraindications to exercise training are included in AHA, ACC, and AACVPR statements and guidelines, which are readily available to practicing clinicians and abstractors. After the condition has been stabilized or reversed, then referral to CR/SP is appropriate. Providers document the specific reason for this exception in clinical notes, summaries and problem lists, which can be abstracted.

2. Health care system factors (e.g., no cardiac rehabilitation program available within 60 minutes of travel time from the patient's home): Although some patients may do so, it is not practical to expect a patient to drive for 2 hours 2 or 3 times per week in order to attend a program that lasts for 1 to 2 hours and research has shown that distance to CR/SP is inversely correlated with attendance We chose 60 minutes (assuming average 30 mph driving speed) based on published data showing that the adjusted odds ratio (OR) to attend CR/SP decreased as the distance from patient zip code to nearest CR/SP facility increased, with the greatest decline between 10.2 (6.5-14.9) miles (OR 0.58) to 31.8 (15.0-231.0) miles (OR 0.29). Although alternative delivery models such as those using telemedicine or home care may be developed in future to provide CR/SP, currently there is no reimbursement for these programs. Therefore, it is unreasonable to hold the provider responsible to refer a patient to a program that he/she is highly unlikely to attend. Providers can determine availability of CR/SP programs from on-line or local resources and document this exception in the medical record. Abstractors can verify the exceptions by cross-referencing the patient's address with publicly available lists of CR/SP program locations.Exclusion:

There is only one exclusion criteria (patients who expired before discharge). This information is readily available within the medical record.

Exceptions:

All eligible patients who can participate in even a low intensity exercise program and who have the cognitive ability to carry out the individualized education and counseling to life-long secondary prevention efforts should be referred to cardiac rehabilitation/secondary prevention programs, because morbidity and mortality benefits extend to nearly all patient populations, regardless of age or co-morbidities. As a result, the exception examples included in the

performance measure relate to either the patient's inability to attend an exercise program (due to physical or practical obstacles) or to cognitive deficits which make them unable to actively participate in exercise or to apply secondary prevention recommendations.

Examples, justification, and data collection issues for exceptions for this measure;

1. Medical factors (e.g., patient deemed by provider to have a medically unstable, life-threatening condition): Medically unstable, life-threatening conditions are contraindications to aerobic exercise and require medical efforts to stabilize and reverse those conditions, rather than efforts directed at secondary prevention of cardiovascular disease. Objective criteria for contraindications to exercise training are included in AHA, ACC, and AACVPR statements and guidelines, which are readily available to practicing clinicians and abstractors. After the condition has been stabilized or reversed, then referral to CR/SP is appropriate. Providers document the specific reason for this exception in clinical notes, summaries and problem lists, which can be abstracted.

2. Health care system factors (e.g., no cardiac rehabilitation program available within 60 minutes of travel time from the patient's home): Although some patients may do so, it is not practical to expect a patient to drive for 2 hours 2 or 3 times per week in order to attend a program that lasts for 1 to 2 hours and research has shown that distance to CR/SP is inversely correlated with attendance We chose 60 minutes (assuming average 30 mph driving speed) based on published data showing that the adjusted odds ratio (OR) to attend CR/SP decreased as the distance from patient zip code to nearest CR/SP facility increased, with the greatest decline between 10.2 (6.5-14.9) miles (OR 0.58) to 31.8 (15.0-231.0) miles (OR 0.29). Although alternative delivery models such as those using telemedicine or home care may be developed in future to provide CR/SP, currently there is no reimbursement for these programs. Therefore, it is unreasonable to hold the provider responsible to refer a patient to a program that he/she is highly unlikely to attend. Providers can determine availability of CR/SP programs from on-line or local resources and document this exception in the medical record. Abstractors can verify the exceptions by cross-referencing the patient's address with publicly available lists of CR/SP program locations.

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

Measure was not stratified. Since all patient sub-groups are reported to have low referral rates and low utilization rates for cardiac rehabilitation services, there is no specific requirement to report data on this performance measure in a stratified format. However, medical centers are encouraged to utilize any stratification of their data as they use the performance measure to identify suboptimal processes and also subgroups at particular risk that are under their care. Such stratification could include stratification by gender, ethnicity, and/or age, since these variables have been found to identify subpopulations that are at particular risk for non-referral to CR/SP in some cities and regions.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

ACC CathPCI Registry calculation:

US HOSP= YES

Discharge date= present

Discharge location=present

Discharge referral= present

Discharge status= present

Exclude any of the below:

-Death

-PCI <= 0

-“NULL” values

ACTION GWTG Registry calculation:

US HOSP= YES

Discharge date= present

Discharge location=present

Discharge referral= present

Discharge status= present

Exclude any of the below:

-Death

-Comfort measure= present

-“NULL” values

AACVPR/ACC/AHA Cardiac Rehabilitation Referral Reliability Testing (CR3) Project:

Hospital ID present = YES

AND

Subject ID = YES

AND

*Provider NPI = YES

AND

Age at start of measurement period is 18 years or older = YES

AND

Qualifying Event: Myocardial Infarction = YES

OR

Qualifying Event: Coronary Artery Bypass Graft = YES

OR

Qualifying Event: Cardiac Valve Surgery = YES

OR

Qualifying Event: Heart Transplantation = YES

OR

Qualifying Event: Stable Angina = YES

OR

Qualifying Event: PCI-stent = YES

OR

Qualifying Event: PCI- other intervention = YES

AND

Yes, documentation that patient was referred to CR for this event/diagnosis

*Since the data for the CR3 Project were processed through the NCDR-PINNACLE Center, NPI was used to help process the data in accordance with the software used at the Center, which requires an NPI on each report. However, since the purpose of the CR3 Project was to assess reliability of the chart abstraction process and not to assess the variability of CR/SP referral by providers, we opted to analyze the CR/SP referral rates by site, and to use the site NPI for data processing purposes only.

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

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

This performance measure is not based on a sample.

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

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

Measure is not based on a survey.

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

If other, please describe in S.18.

Electronic Health Records, Registry Data

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

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

American College of Cardiology PINNACLE registry and AACVPR/ACC/AHA Cardiac Rehabilitation Testing (CR3) Project.

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

Available in attached appendix at A.1

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

Clinician : Individual, Facility, Integrated Delivery System

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

Inpatient/Hospital

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

NQF_testing_attachment_Sep2017_0642.pdf

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the

most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.2 For maintenance of endorsement

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

No

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** 0642
Measure Title: Cardiac Rehabilitation Patient Referral from an Inpatient Setting
Date of Submission: 11/8/2017

Type of Measure:

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

Instructions

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

- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.

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

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

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

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

AND

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

Notes

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

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face

validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

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

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

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

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

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

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

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

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

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

The following datasets were used: AACVPR/ACCF/AHA Cardiac Rehabilitation Referral and Reliability (CR3) Project and the ACCF/AHA ACTION-GWTG Registry

1.3. What are the dates of the data used in testing? August 2009-December 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.20***)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☒ group/practice	☒ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ 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*)

AACVPR/ACCF/AHA CR3 Project:U.S. hospitals identified from the ACCF, AHA, and AACVPR databases were invited to participate in the Cardiac Rehabilitation Referral and Reliability (CR3) Project. We sought a variety of hospitals, based on varied geographical locations, community sizes, and hospital types/sizes. Hospitals that met participation criteria were included in the project. Participation criteria included a willingness and ability to: (1) provide a study coordinator and 2 chart abstractors, (2) complete the project within the specified timeline, and (3) obtain local IRB approval to carry out the project in their hospital. Once each hospital completed and submitted their required data, they were sent a small token of appreciation from AACVPR, ACCF, and AHA. A total of 45 hospitals expressed an interest in participating in the project, including hospitals from outside the U.S. (Puerto Rico, Romania, and Turkey). 7 hospitals (all in the United states and distributed around the country) met all participation criteria and were selected to participate in the project. The sites used a mixture of paper medical records and EHR systems.

ACCF/AHA ACTION-GWTG Registry:

The ACTION-GWTG Registry is an in-patient registry of ACCF and AHA that collects information on patients who have been admitted to a participating in-patient center due to a myocardial infarction. A number of clinical characteristics are collected, including patient characteristics, treatments, and discharge orders/plans.

For purposes of reporting the scientific testing that we have carried out on the Performance Measure for Cardiac Rehabilitation Referral from an Inpatient Facility, we have included data analyses from 3 cohorts from the ACTION-GWTG Registry, 2012: Data were analyzed on 122,285 patients from 703 in-patient centers in the ACTION-GWTG Registry.

ACCF CathPCI Registry:

Testing results are included from 623,098 patients from 1371 inpatient centers participating in the ACCF CathPCI Registry from January 1-December 31, 2012. Trends in cardiac rehabilitation referral were analyzed for the 2012 cohort, and also for 616,545 patients included in the registry from January 1, 2011 through December 31, 2011.

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*)

AACVPR/ACCF/AHA CR3 Project:

Descriptive statistics are noted below:

Sex: Male: 65% (n=152 / 234), Female: 35% (n=82/234)

Age: 18-39: 3% (n=7/229), 40-64: 40% (n=91 / 229), 65-79: 45% (n=103 / 229) , 80+: 12% (n=28 / 229)

Race: White: 84% (n=196 / 234), Black: 8% (n=19 / 234), Asian: 0% (n=1 / 234), American

Indian: 1% (n=3 / 234), Native Hawaiian/Pacific Islander: 0% (n=1 / 234), Other: 6% (n=14 /234)

Hispanic Ethnicity: 0% (n=1 / 234)

ACTION-GWTG Registry, 2012:

CR

	Total	Cardiac Rehab Referral		P-Value
	n = 122285	Refer n = 92362	Not Refer n = 29923	
age	63.3 ± 13.4	62.7 ± 13.1	65.0 ± 14.1	< 0.001
Male Gender	81201 (66.4%)	62725 (67.9%)	18476 (61.7%)	< 0.001
racecat				
1 Caucasian	103641 (84.8%)	79246 (85.8%)	24395 (81.5%)	< 0.001
2 Af Am	14329 (11.7%)	10308 (11.2%)	4021 (13.4%)	
3 Other	4315 (3.5%)	2808 (3.0%)	1507 (5.0%)	
inscat				
1 Private	70170 (57.4%)	54457 (59.0%)	15713 (52.5%)	< 0.001
2 Medicare	28803 (23.6%)	20205 (21.9%)	8598 (28.7%)	
3 Medicaid	5273 (4.3%)	3713 (4.0%)	1560 (5.2%)	
4 Other	2949 (2.4%)	2192 (2.4%)	757 (2.5%)	
5 None	15090 (12.3%)	11795 (12.8%)	3295 (11.0%)	
smoker	44483 (36.4%)	34869 (37.8%)	9614 (32.1%)	< 0.001
Missing (.)	29	24	5	
Prior PAD	10471 (8.6%)	7561 (8.2%)	2910 (9.7%)	< 0.001
Missing (.)	89	58	31	
Prior CVD	12809 (10.5%)	9090 (9.8%)	3719 (12.4%)	< 0.001
Missing (.)	59	34	25	
Prior PCI	24470 (25.1%)	18713 (24.6%)	5757 (26.7%)	< 0.001
Missing (.)	24776	16442	8334	
Prior MI	23881 (24.5%)	18021 (23.7%)	5860 (27.1%)	< 0.001
Missing (.)	24773	16443	8330	
Prior HF	10256 (10.5%)	6955 (9.2%)	3301 (15.3%)	< 0.001
Missing (.)	24908	16546	8362	
Prior CABG	12791 (13.1%)	9378 (12.4%)	3413 (15.8%)	< 0.001
Missing (.)	24788	16453	8335	
Currently on Dialysis	2603 (2.1%)	1620 (1.8%)	983 (3.3%)	< 0.001
Missing (.)	121	68	53	
Hypertension	87317 (73.1%)	65575 (72.1%)	21742 (76.5%)	< 0.001
Missing (.)	39	23	16	
Diabetes	38500 (32.3%)	28390 (31.2%)	10110 (35.6%)	< 0.001
Missing (.)	71	41	30	

Continuous variables compared using Student's T-test.

Categorical variables compared using chi-square or Fisher's exact test.

ACCF CathPCI Registry, 2012:

Rehab

	Total	Cardiac Rehabilitation Referral		P-Value
	n = 623098	Yes n = 383112	No n = 239986	
History				
Age	64.6 ± 12.0	64.3 ± 12.0	65.1 ± 12.0	< 0.001
Sex				< 0.001
Male	424459 (68.1%)	261946 (68.4%)	162513 (67.7%)	
Female	198639 (31.9%)	121166 (31.6%)	77473 (32.3%)	
IABP	12198 (2.0%)	7705 (2.0%)	4493 (1.9%)	< 0.001
Missing (.)	164	85	79	
Current/Recent Smoker (w/in 1 year)	172783 (27.7%)	110266 (28.8%)	62517 (26.1%)	< 0.001
Missing (.)	351	195	156	
Hypertension	512238 (82.2%)	311186 (81.3%)	201052 (83.8%)	< 0.001
Missing (.)	199	129	70	
Dyslipidemia	489637 (78.7%)	299362 (78.2%)	190275 (79.4%)	< 0.001
Missing (.)	595	373	222	
Family History of Premature CAD	155296 (24.9%)	96057 (25.1%)	59239 (24.7%)	< 0.001
Missing (.)	246	141	105	
Prior MI	188626 (30.3%)	114869 (30.0%)	73757 (30.7%)	< 0.001
Missing (.)	160	80	80	
Prior Heart Failure	74910 (12.0%)	44360 (11.6%)	30550 (12.7%)	< 0.001
Missing (.)	271	180	91	
Prior Valve Surgery/Procedure	9336 (1.5%)	5403 (1.4%)	3933 (1.6%)	< 0.001
Missing (.)	339	212	127	
Prior PCI	253945 (40.8%)	152328 (39.8%)	101617 (42.4%)	< 0.001
Missing (.)	154	68	86	
Prior CABG	111609 (17.9%)	67268 (17.6%)	44341 (18.5%)	< 0.001
Missing (.)	99	55	44	
Currently on Dialysis	14746 (2.4%)	7698 (2.0%)	7048 (2.9%)	< 0.001
Missing (.)	578	354	224	
Cerebrovascular Disease	76660 (12.3%)	46559 (12.2%)	30101 (12.5%)	< 0.001
Missing (.)	267	174	93	

	Total	Cardiac Rehabilitation Referral		P-Value
	n = 623098	Yes n = 383112	No n = 239986	
Peripheral Arterial Disease	76367 (12.3%)	45187 (11.8%)	31180 (13.0%)	< 0.001
Missing (.)	267	175	92	
Chronic Lung Disease	93876 (15.1%)	57218 (14.9%)	36658 (15.3%)	< 0.001
Missing (.)	269	181	88	
Diabetes Mellitus	231186 (37.1%)	138108 (36.1%)	93078 (38.8%)	< 0.001
Missing (.)	300	97	203	
Cath Lab Visit				
PCI Indication				< 0.001
Immediate PCI for STEMI	91297 (14.7%)	63260 (16.5%)	28037 (11.7%)	
PCI for STEMI (Unstable, >12 hrs from Sx onset)	5512 (0.9%)	3630 (0.9%)	1882 (0.8%)	
PCI for STEMI (Stable, >12 hrs from Sx onset)	2621 (0.4%)	1672 (0.4%)	949 (0.4%)	
PCI for STEMI (Stable after successful full-dose Thrombolysis)	2129 (0.3%)	1481 (0.4%)	648 (0.3%)	
Rescue PCI for STEMI (after failed full-dose lytics)	3115 (0.5%)	2308 (0.6%)	807 (0.3%)	
PCI for high risk Non-STEMI or unstable angina	324113 (52.0%)	203550 (53.1%)	120563 (50.3%)	
Staged PCI	43430 (7.0%)	24502 (6.4%)	18928 (7.9%)	
Other	150724 (24.2%)	82636 (21.6%)	68088 (28.4%)	
Missing (.)	157	73	84	
CAD Presentation				< 0.001
No symptom, no angina	38290 (6.1%)	21232 (5.5%)	17058 (7.1%)	
Symptom unlikely to be ischemic	13990 (2.2%)	7734 (2.0%)	6256 (2.6%)	
Stable angina	89099 (14.3%)	49158 (12.8%)	39941 (16.7%)	
Unstable angina	249446 (40.0%)	149336 (39.0%)	100110 (41.7%)	
Non-STEMI	129825 (20.8%)	84659 (22.1%)	45166 (18.8%)	
ST-Elevation MI (STEMI) or equivalent	102284 (16.4%)	70931 (18.5%)	31353 (13.1%)	
Missing (.)	164	62	102	
Anginal Classification w/in 2 Weeks				< 0.001
No symptoms	58945 (9.5%)	32652 (8.5%)	26293 (11.0%)	

	Total	Cardiac Rehabilitation Referral		P-Value
	n = 623098	Yes n = 383112	No n = 239986	
CCS I	22585 (3.6%)	11160 (2.9%)	11425 (4.8%)	
CCS II	90921 (14.6%)	49037 (12.8%)	41884 (17.5%)	
CCS III	226193 (36.3%)	140557 (36.7%)	85636 (35.7%)	
CCS IV	223642 (35.9%)	149273 (39.0%)	74369 (31.0%)	
Missing (.)	812	433	379	
Anti-Anginal Medication w/in 2 Weeks	450685 (72.4%)	276280 (72.1%)	174405 (72.7%)	< 0.001
Missing (.)	187	110	77	
Heart Failure w/in 2 Weeks	62229 (10.0%)	37442 (9.8%)	24787 (10.3%)	< 0.001
Missing (.)	264	135	129	
Cardiomyopathy or Left Ventricular Systolic Dysfunction	65458 (10.5%)	40176 (10.5%)	25282 (10.5%)	0.544
Missing (.)	150	87	63	
Pre-operative Evaluation Before Non-Cardiac Surgery	11296 (1.8%)	6354 (1.7%)	4942 (2.1%)	< 0.001
Missing (.)	214	121	93	
Cardiogenic Shock w/in 24 Hours	8729 (1.4%)	5608 (1.5%)	3121 (1.3%)	< 0.001
Missing (.)	110	63	47	
Cardiac Arrest w/in 24 Hours	10045 (1.6%)	6685 (1.7%)	3360 (1.4%)	< 0.001
Missing (.)	180	101	79	
Pre-PCI Left Ventricular Ejection Fraction	52.5 ± 12.3	52.5 ± 12.2	52.6 ± 12.5	0.012
Missing	183357	113926	69431	
Procedure Information				
Contrast Volume	190.6 ± 87.3	192.0 ± 86.7	188.4 ± 88.2	< 0.001
Missing	1680	966	714	
Fluoroscopy Time	14.8 ± 11.6	14.6 ± 11.5	15.1 ± 11.8	< 0.001
Missing	8457	5441	3016	
Outcomes				
Myocardial Infarction (Biomarker Positive)	12321 (2.0%)	7092 (1.9%)	5229 (2.2%)	< 0.001
Missing (.)	195	122	73	
Cardiogenic Shock	4560 (0.7%)	2826 (0.7%)	1734 (0.7%)	0.496
Missing (.)	184	113	71	

	Total	Cardiac Rehabilitation Referral		P-Value
	n = 623098	Yes n = 383112	No n = 239986	
Heart Failure	5795 (0.9%)	3673 (1.0%)	2122 (0.9%)	0.003
Missing (.)	191	118	73	
CVA/Stroke	1079 (0.2%)	656 (0.2%)	423 (0.2%)	0.642
Missing (.)	196	122	74	
Other Vascular Complications Requiring Treatment	2357 (0.4%)	1450 (0.4%)	907 (0.4%)	0.972
Missing (.)	200	127	73	
RBC/Whole Blood Transfusion	12607 (2.0%)	7824 (2.0%)	4783 (2.0%)	0.180
Missing (.)	200	121	79	

Continuous variables compared using Student's T-test.

Categorical variables compared using chi-square or Fisher's exact test.

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 described above was used for all aspects of testing

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

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*)

AACVPR/ACCF/AHA CR3 Project:

> The aim of this project was to assess the reliability and feasibility of abstracting the Cardiac Rehabilitation Referral Performance Measure from an inpatient setting. The sites identified to participate in the project were asked to identify one study coordinator and two chart abstractors. 35 patients were identified by the study coordinator at each site from a consecutive sample of patients admitted to their hospital having a qualifying diagnosis for CR, and discharged alive, starting in August 1, 2009. The two abstractors at each site reviewed the same 35 patient records from their hospital twice (once at baseline, and again one week later).

Site coordinators were instructed to include in the 35 patient records, 30 patients who had qualifying diagnoses for CR/SP referral (to capture sensitivity testing) and 5 patients who did not have a qualifying diagnosis for CR/SP

referral (to capture specificity testing). The qualifying events are indicated in the measure numerator statement. The non-qualifying events for the purpose of this abstraction project needed to have one or more of the following diagnoses: heart failure, atrial fibrillation, or syncope.

The CR3 Project Workgroup worked with the study coordinators to address reliability, feasibility, and usability properties for the cardiac rehabilitation performance measures. Specifically the workgroup created chart abstraction forms, site coordinator instructions, abstractor instructions, sample IRB protocol, frequently asked questions, and tracking forms to keep track of the intra-rater (1 abstractor reviewing the same patient record two times) and inter-rater process (2 abstractors reviewing the same patient record). The workgroup had a kickoff call with each center's study coordinator to acquaint him/her with the abstraction project. The workgroup communicated weekly with site coordinators to address any questions or comments the sites may have had.

Abstractors reviewed each patient record and completed the CR3 Project form (see supplement). Definitions used:

Eligible patient: a patient that had a qualifying event/diagnosis during the hospital period under review

Non-eligible patient: a patient that did not have a qualifying event/diagnosis during the hospital period under review

CR/SP referral : documentation in the patient record for the index hospitalization that the patient was being referred to an outpatient cardiac rehabilitation/secondary prevention program

Exception to referral : documentation in the patient record for the index hospitalization that a patient who was eligible for CR/SP referral had a patient, medical, or healthcare system exception that prohibited their participation in CR/SP

Analyses were performed as follows:

1. Intra-rater and inter-rater agreement between patient record reviews Eligibility:
   Was the patient eligible for CR/SP referral?

   CR/SP Referral: Was each eligible patient referred to CR/SP?

   Exceptions: For patients not referred to CR/SP, was/were any exception(s) to CR/SP documented?
2. Percent agreement

   In what percentage of patient record abstractions did the abstractors agree (for both intra-rater and inter-rater agreement)?
3. Kappa statistic

Site specific: Calculated for the 2 abstractors at each site, to compare intra- and inter-abstractor reliability, with regards to his/her assessment of: (1) eligibility for CR/SP referral, (2) referral to CR/SP, and (3) exceptions to CR/SP referral

Pooled estimate: data from all sites were combined to calculate a pooled kappa statistic for intra- and inter-observer reliability for assessing CR/SP eligibility, referral, and exceptions.

By convention, a kappa > .70 is considered acceptable inter-rater reliability.(1) 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.)

It is important to consider both the "percent agreement" and the kappa statistic when assessing the reliability of abstracting this performance measure from patient records, especially for the assessment of "eligibility" and "exceptions". Each method of reliability assessment gives a slightly different view of reliability in this case.

"Percent agreement" is a helpful assessment of reliability of the measure, but given that over 80% of the patients in the study sample were eligible for cardiac rehabilitation, and more than 90% of the patients were free from exceptions to cardiac rehabilitation participation, the percent agreement for the abstractors may have been somewhat inflated, since by chance alone abstractors may have chosen the "right" eligibility or exception status. (To help minimize this, we blinded the abstractors to the actual number/percentage of patients who were eligible for cardiac rehabilitation in their sample. In addition, abstractors were unaware of the range of exceptions that would be expected in their sample.)

The kappa statistic performs best when there is nearly equal chance of study outcomes (for example, equal chance of being eligible or not eligible for cardiac rehabilitation). When there is a high likelihood of one of the two outcomes, as in our study (high likelihood of eligibility), the results of the kappa analyses can sometimes be less accurate and actually underestimate the true reliability the measure due to a phenomenon that is referred to as a "kappa score paradox" in which there is high percent agreement, yet a low kappa score. (Reference: Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. *J Clin Epidemiol.* 1996 Apr;49(4):431-4.) Indeed, we observed in our site specific analyses that in some centers with very high percent agreement within and between abstractors, the kappa statistics were very low or even zero in some rare cases. With this in mind, the kappa statistic may underestimate the true reliability of the CR measure.

Using both the "percent agreement" and the kappa statistic together provides a robust view of the reliability of the CR performance measure. One ("percent agreement") may slightly overestimate reliability and the other (kappa statistic) may slightly underestimate reliability. The true reliability of the measure most likely lies between the results from the two methods of assessment. Since the "percent agreement" method suggests "almost perfect" reliability and the kappa statistic suggests "substantial" to "almost perfect" reliability, the overall reliability of the CR performance measure appears to be between "substantial" and "almost perfect"

ACTION-GWTG Registry:

Data were used to assess reliability and other performance characteristics for centers participating in the ACTION-GWTG Registry from January 1 2011 until December 31, 2012.

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 five different points: 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.

ACCF CathPCI Registry:

Data were used to assess reliability and other performance characteristics for centers participating in the ACCF CathPCI Registry from January 1, 2011 until December 31, 2012. Reliability was analyzed by calculating signal-

to-noise ratios, using the same methods described above.

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*)

AACVPR/ACCF/AHA CR3 Project (pooled data results):

The abstractor and coordinator experiences in chart abstraction prior to participating in the CR3 project varied greatly. The summary data is below.

Less than 1 month 39% (n= 11)

1-6 months 11% (n=3)

6-12 months 7% (n=2)

1-2 yrs 4% (n=1)

2-3 yrs 4% (n=1)

3-4 yrs 11% (n=3)

4-5 yrs___None

More than 5 years 25% (n=7)

ARE PATIENTS ELIGIBLE FOR CARDIAC REHABILITATION?

Percentage deemed eligible for cardiac rehabilitation: 199 / 234 (85%) (mean of all observations) (Actual percentage of patients who were eligible for cardiac rehabilitation: 200/234 (86%))

Intra-rater reliability (agreement within the same abstractor):

% Agreement: 232 / 232 (100%)

Kappa: 1.00 (-)

Inter-rater reliability (agreement between abstractors):

% Agreement: 218 / 231 (94%)

Kappa: 0.77 (0.65, 0.89)

HAVE PATIENTS BEEN REFERRED TO CARDIAC REHABILITATION?

Percentage referred to cardiac rehabilitation: 111 /185 (60%) (mean of all observations)

Intra-rater reliability:

% Agreement: 172 / 176 (98%)

Kappa: 0.95 (0.90, 0.99)

Inter-rater reliability:

% Agreement: 148 / 172 (86%)

Kappa: 0.70 (0.59, 0.81)

ARE THERE EXCEPTIONS NOTED FOR ELIGIBLE PATIENTS NOT REFERRED TO CARDIAC REHABILITATION?

Percentage with documented exceptions to cardiac rehabilitation: 17 /201 (9%) (mean of all observations)

Intra-rater reliability:

% Agreement: 189 / 196 (96%)

Kappa: 0.76 (0.60, 0.93)

Inter-rater reliability:

% Agreement: 185 / 191 (97%)

Kappa: 0.79 (0.63, 0.95)

ACTION-GWTG Registry, 2012:

Results of reliability testing (signal-to-noise ratios) are as follows:

Level	Signal-to-Noise
All, >10 Procedures	.988
>Q1	.993
>Q2	.995
>Q3	.997
>Average	.996

ACCF CathPCI Registry, 2012:

Results of reliability testing (signal-to-noise ratios) are shown below:

Level	Signal-to-Noise
All, >10 Procedures	.996
>Q1	.998
>Q2	.999
>Q3	.999
>Average	.999

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?*)

AACVPR/ACCF/AHA CR3 Project:

The percentage agreement within and between abstractors was “almost perfect” for eligibility, referral and exceptions. The kappa statistic for agreement within and between abstractors was “substantial” to “almost perfect” for referral, and “moderate” for exceptions. Given the very low exception rates, it is possible that the kappa statistic underestimates reliability of abstracting the measure, due to the phenomenon of a kappa “paradox” (lower kappa statistic despite high percent agreement within and between abstractors, related to the low “event” or exception rates). (lower kappa statistic despite high percent agreement within and between abstractors, related to the low “event” or exception rates, see item 3 in the analytic method section above for more details). This conclusion is even more likely given that we asked site coordinators to identify 30 of 35 patient records who presented with an eligible diagnosis/event.

ACTION-GWTG Registry, 2012:

Results of the reliability testing show that the measure has excellent reliability when evaluated at the minimum level of quality reporting events, and higher reliability at the median number of events (50^{th}%), and at average and greater number of quality events.

ACCF CathPCI Registry, 2012:

Results of the reliability testing show that the measure has excellent reliability when evaluated at the minimum level of quality reporting events, and higher reliability at the median number of events (50^{th}%), and at average and greater number of quality events

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☐ **Empirical validity testing- Will aim to obtain additional empirical validity testing data for future measures as time allows**

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

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

Face and Content Validity

Validity Survey of Experts: Validity of the measure score was systematically assessed as follows: After the measure was fully specified, members of 3 existing committees, one at the ACC, one at AHA and one joint ACC/AHA, with expertise in general cardiology, cardiac rehabilitation, quality improvement, outcomes research, 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

Face and content validity of the measure score was systematically assessed according to responses received from survey respondents.

AACVPR/ACCF/AHA performance measures set: To determine the content/context validity of the measures, a process using a Delphi peer review was utilized. An explicit and standardized process for ACCF/AHA performance measure development was followed, including the following steps: 1.

Formation of the Development Committee, 2. Identification of Potential Factors, for Inclusion, 3. Scoring of the Factors/Expert Opinion, 4. Public Comment Period/Peer Review, 5. Further Refinement, 6. Final approval by organizations, 7. Peer Review Publication/Endorsement. Reviewers were asked to provide comments on the document on the basis of the rating form and guide shown on page 1432 at Http://content.onlinejacc.org/cgi/reprint/j.jacc.2007.04.033v1.pdf

Content/context validity of the measures was also established by virtue of the specialized expertise of the Performance Measures Work Group members including the structured discussions that the work group conducted, as well as the rigorous peer review and public comment period that were carried out. For this particular topic those individuals who were involved in identifying and drafting the performance measures were leaders and experts in the field of cardiac rehabilitation as chosen by AACVPR, ACCF, and AHA.

Furthermore, additional face and content validity was demonstrated from the update of the measure in 2010. During the NQF Care Coordination project, the Steering Committee asked AACVPR, ACCF, and AHA to remove patient refusal as an exception. Since that time, all 3 organizations have published an updated document (NQF measures 0642 and 0643) that explicitly notes that patient refusal should not be an allowable exception. In addition, the cardiac rehabilitation referral measures were revised to facilitate the implementation of these two measures by including administrative codes to identify denominator-eligible populations. All changes were approved by the American Association of Cardiovascular and Pulmonary Rehabilitation Board of Directors, the American College of Cardiology Foundation Board of Trustees, and by the American Heart Association Science Advisory and Coordinating Committee.

The performance measure set was also reviewed via AHA and ACC processes as well as by the AACVPR Document Oversight Committee.

AACVPR/ACCF/AHA CR3 Project: Through the NQF endorsement process, the cardiac rehabilitation referral performance measures ("Set A" measures) received time-endorsed status in 2010, thus supporting the content validity of these measures.

ACTION-GWTG Registry and ACCF CathPCI Registry: ACCF and AHA registries always attempt to include ACCF/AHA Task Force on Performance Measures in their various modules. The measures have content/context validity based on the approach articulated under the AACVPR/ACCF/AHA performance measure set.

Predictive Validity: Published data have shown that as the cardiac rehabilitation referral measure is met (i.e., the patient is referred to cardiac rehabilitation), the proximate desired outcome (cardiac rehabilitation participation) increases, as does the longer term desired outcome (reduction in morbidity and mortality rates). For more details, see the supplemental materials.

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

Face Validity Survey of Experts:

There were 17 individuals who completed the survey. Further information on the survey respondents is available if needed. Results of the survey were as follows:

-Average score: 4.18

-88.24% of respondents either agree or strongly agree that the outpatient measure can accurately distinguish good and poor quality.

AACVPR/ACCF/AHA performance measures set:

In May 2007 the final peer reviewed publication of the performance measures document was approved by the American Association of Cardiovascular and Pulmonary Rehabilitation Board of Directors, the American College of Cardiology Foundation Board of Trustees and by the American Heart Association Science Advisory and Coordinating Committee. Additionally, the publication was endorsed by the American College of Chest Physicians, American College of Sports Medicine, American Physical Therapy Association, Canadian Association of Cardiac Rehabilitation, European Association for Cardiovascular Prevention and Rehabilitation, Inter-American Heart Foundation, National Association of Clinical Nurse Specialists, Preventive Cardiovascular Nurses Association, and the Society of Thoracic Surgeons. The final document was published *the Journal of the American College of Cardiology* (the official journal of the American College of Cardiology), *the Journal of Cardiopulmonary Rehabilitation and Prevention* (the official journal of the American Association of Cardiovascular and Pulmonary Rehabilitation) and *Circulation* (the official journal of the American Heart Association) in September 2007. The document can be found at http://content.onlinejacc.org/cgi/reprint/j.jacc.2007.04.033v1.pdf.

AACVPR/ACCF/AHA CR3Project:

The cardiac rehabilitation referral measures (NQF measures 0642 and 0643) were revised in 2010 to clarify numerator and denominator exclusion criteria and to facilitate the implementation of these two measures by including administrative codes to identify denominator-eligible populations.

ACTION-GWTG Registry and ACCF CathPCI Registry:

A review of the measure based on the attributes, of reliability, ease of implementation, appropriate numerator, denominator, and exception specifications was performed. Given that it fulfilled these attributes, the measure was included in the registry. Data from these registries can be seen throughout the submission form and supplemental materials.

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

As noted above, our interpretation is that face and content validity has been established for this measure.

_____

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section* 2b4

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

AACVPR/ACCF/AHA CR3Project: Reliability of abstracting measure exclusions was tested in the CR3 project (see reliability testing section above). Exclusions or exceptions include patient, medical, and system-based conditions that would preclude the reasonable participation of a patient in a cardiac rehabilitation program (death, residing in an extended care nursing facility, lack of a cardiac rehabilitation program close to where the patient lives, etc.).

ACTION-GWTG Registry and ACCF CathPCI Registry, 2012: Exclusion rates and reasons were assessed from the ACTION-GWTG and CathPCI Registries.

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

The reliability of abstracting exclusions was high to very high, as shown in the reliability section of this document.

In the ACTION-GWTG Registry, the exclusion rate was 13.6% (18,792/138/201). Out of 665 hospitals included, 80 had no exclusions and the remaining 585 had a mean exclusion rate of 16.6%.

In the CathPCI Registry, the exclusion rate was 4.3% (27,830/650,928). Out of 1360 hospitals included, 955 had no exclusions and the remaining 405 hospitals had a mean exclusion rate of 6.4%.

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

Documentation and assessment of exclusions are very important for this measure, to help reduce the possibility of bias in reporting (i.e., excluding patients who are actually eligible for cardiac rehabilitation referral, in order to improve performance scores). Based on he results of our CR3 project, the time and effort to assess exclusions does not appear to add significant burden (see supplemental materials section for more details on time required to complete abstraction of data for this measure).

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

***If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section* 2b5.**

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

☒ No risk adjustment or stratification

☐ Statistical risk model with _risk factors

☐ Stratification by _risk categories

☐ Other,

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

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

Referral to cardiac rehabilitation is appropriate and evidence-based for all patients who have had a qualifying

event/diagnosis/procedure. Referral or non-referral is not based on a patient's level of risk, but rather cardiac rehabilitation is appropriate and evidence-based for all eligible patients no matter what their risk level.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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

If stratified, skip to 2b3.9

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

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

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

Performance rates were assessed by chart abstraction in the CR3 project. Reliability of that assessment was also performed, as noted in the reliability section above.

In the ACTION-GWTG and CathPCI Registries, performance rates were assessed by decile, to allow for assessment of differences between "low" and "high" performing centers.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

2012 (ACTION Registry-GWTG)

Analysis Variable : P Proportion Referral							
Number Hospitals	Mean	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Quartile Range
703	0.6698336	0	0.4657763	0.8014184	0.9456522	1.0000000	0.4798759

By Decile:

10th Percentile	20th Percentile	30th Percentile	40th Percentile	50th Percentile	60th Percentile	70th Percentile	80th Percentile	90th Percentile
0.065217	0.31429	0.59091	0.69369	0.80142	0.87302	0.92537	0.95876	0.98701

2012 (CathPCI)

Analysis Variable : P Proportion CR Reffer							
N	Mean	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Quartile Range
1360	0.5936149	0	0.1774152	0.7233546	0.9422948	1.0000000	0.7648796

By Decile:

10th Percentile	20th Percentile	30th Percentile	40th Percentile	50th Percentile	60th Percentile	70th Percentile	80th Percentile	90th Percentile
0.019311	0.097453	0.33059	0.56213	0.72335	0.84518	0.91913	0.95949	0.98632

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

There is wide variation in performance for this measure, documented in the datasets we used. Use of this measure allows for identification of that variation in delivery of cardiac rehabilitation referral. This is important because it provides data from which centers can identify improve upon gaps in care that are identified.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

In the CR3 Project, lack of documentation of a cardiac rehabilitation referral was assumed to represent "no referral made". In the ACTION and CathPCI databases, a missing response was identified when there was no response to the cardiac rehabilitation measure.

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

Given our assumptions, noted above, we did not conduct an empirical analysis of the frequency or distribution of missing data in the CR3 project. For this measure, missing data represents a failure. In the ACTION dataset, the missing data rate for our primary variable was extremely low at 0.58% (n=810). In the CathPCI dataset, the missing data rate for our primary variable was extremely low at 0.14% (n=931).

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing*

data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data)

Our assumption, based on the data listed above, is that the missing data rate is extremely low for our primary 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.

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

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

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

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

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

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

Attachment:

3c. Data Collection Strategy

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

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

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

ACC CathPCI Registry:

Referral to CR/SP following percutaneous intervention is low, compared to other performance measures, as documented by the ACC CathPCI Registry. The work group recognizes that this is due to several factors, some of which can be modified with education and improved processes. Many cardiologists remain confused about insurance coverage for CR/SP following percutaneous intervention, because Medicare did not cover these services until after legislative changes in 2006. AACVPR has been working with the Society for Cardiovascular Angiography and Intervention (SCAI) and ACC to provide educational materials for cardiologists and their patients about the benefits of CR/SP. These include fact sheets for patients with a space to insert specific program contact information, both in English and Spanish, and enhanced on-line educational materials.

ACTION Registry:

Based on our observation that inter-rater reliability testing with this registry was not as strong as compared to the CR3 project, we suspect that additional education related to implementation of the CR/SP referral measures is needed. The measure testing workgroup plans to work with AACVPR, ACC, and AHA leadership to develop implementation notes to instruct abstractors and providers about the documentation details needed to meet the 3 components of CR/SP referral criteria.

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 $995-$15,000.

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.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Payment Program	Professional Certification or Recognition Program ACTION Registry Achievement Award https://cvquality.acc.org/NCDR-Home/registries/hospital-registries/action-registry/action-registry-performance-achievement-awards Quality Improvement (external benchmarking to organizations) NCDR CathPCI registry https://www.ncdr.com/webncdr/cathpci/ NCDR ACTION Registry https://www.ncdr.com/webncdr/action/ Quality Improvement (Internal to the specific organization) ACC Patient Navigator https://cvquality.acc.org/initiatives/patient-navigator

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

NCDR CathPCI Registry:

The CathPCI Registry is sponsored by ACC in conjunction with the Society for Cardiovascular Angiographyand Interventions. The CathPCI Registry was designed to create a national surveillance system to assess the characteristics, treatments, and outcomes of patients with coronary heart disease who undergo procedures in cardiac catheterization laboratories. Eligible patients are adults (18 years of age and older) who undergo a diagnostic cardiac catheterization and/or PCI. More than 1,300 hospitals across the U.S submit data to the CathPCI registry. Participation in the CathPCI Registry provides risk-adjusted quarterly benchmark reports that compares an institution's performance with that of volume-based peer groups and the national experience. The registry includes standardized, evidence-based data elements and definitions, a Dashboard tool that provides a custom query to control for variables (facility size, number of procedures, teaching vs. non-teaching sites, states and regions) to compare the participating facility data, metrics and volumes. ABIM Diplomates can also meet MOC recertification requirements by using CathPCI Registry data to earn up to 80 points toward evaluation of practice performance through the Clinical Quality Coach mobile app.

The NCDR ACTION Registry

The ACTION (Acute Coronary Treatment and Intervention Outcomes Network Registry) is sponsored by the ACCF, with partnering support from The American College of Emergency Physicians and The Society of Hospital Medicine. The ACTION Registry was designed to assess the characteristics, treatments, and outcomes of acute myocardial infarction (AMI) patients (either ST-segment elevation myocardial infarction or non–ST-segment elevation myocardial infarction). Eligible patients are those older than 18 years of age hospitalized with a diagnosis of AMI who have acute ischemic symptoms within 24 h of presentation. Patients admitted for other conditions who subsequently develop AMI during hospital stay are not included. More than 900 hospital across the U.S. submit data to ACTION Registry.

ACTION Registry Achievement Award:

For the demonstration of achievement by sustaining performance measures in the treatment of acute myocardial infarction patients through the implementation of ACTION Registry® and in-hospital initiation of the American College of Cardiology/American Heart Association Clinical Guideline recommendations.

ACC Patient Navigator

The ACC has launched a national scale program, the Patient Navigator Program: Focus MI, to improve the care and outcomes of myocardial infarction patients and further reduce avoidable readmissions beyond 30 days. The ACC ACTION registry is a part of this program.

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

ACC plans on reporting this measure publicly through the CathPCI QCDR. In addition, the ACC is expanding public reporting for the ACTION registry. While the ACC has not yet finalized which measures are moving forward for public reporting, NQF-endorsed measures are highly desirable. Lastly, the ACC is also in active discussions with a private insurance payer entity. Due to non- disclosure requirements at this time, we are not able to provide specific details. We anticipate being able to provide more details by the January CV project meeting.

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 AACVPR 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.

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

The ACC has also made a decision to voluntarily public report out of the ACTION registry. The data release consent release from for ACTION is available on the CV quality website: https://cvquality.acc.org/docs/default-source/ncdr/public-reporting-documents/action-registry-public-reporting-v2.

ACC's National Cardiovascular Data Registry (NCDR) Voluntary Hospital Public Reporting Program: The ACC currently runs a program to give hospitals the opportunity to voluntarily publicly report their measure results based on data from the National Cardiovascular Data Registry (NCDR). Hospitals that choose to participate have their results displayed on ACC's CardioSmart. Currently Hospitals can report on five measures from the CathPCI Registry and five measures from the ICD Registry. Of these publicly reporting measures, five are NQF-endorsed:

NQF # 1522: Use of a medicine in the ACEi or ARB class to improve heart function after ICD implant in patients with less than normal heart function.

NQF # 1528: Use of a beta-blocker medication after ICD implant in patients with a previous heart attack.

NQF #1529: Use of a beta-blocker medication after ICD implant in patients with less than normal heart function.

NQF #0965: Use of all recommended medications (ACEI or ARB and beta-blocker) to improve heart function and blood pressure after ICD implant.

NQF # 0964: Therapy with aspirin, P2Y12 inhibitor, and statin at discharge following PCI in eligible patients (composite measure)

Starting in 2018, participants can also elect to report on ACTION Registry metrics of which CR referral from an inpatient setting is a part of.

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

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

Performance results are distributed to all CathPCI and ACTION registry participants as part of quarterly benchmark reports, which provide a detailed analysis of an institution´s individual performance in comparison to the entire registry population from participating hospitals across the nation. Reports include an executive summary dashboard, at-a-glance assessments, and patient level drill-downs. Registry participants also have access to an outcome report companion guide which provides common definitions and detailed metric specifications to assist with interpretation of performance rates.

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

Results are provided as part of quarterly performance report which includes a rolling 4 quarters of data.

Participating hospitals in the CathPCI registry report the following: Patient demographics for cardiac catheterization and PCI procedures, provider and facility characteristics, history/factors, cardiac status, treat lesions; intracoronary device utilization and adverse event rates; appropriate use criteria for coronary revascularization; compliance with ACC/AHA clinical guideline recommendations.

Participating hospitals in the ACTION registry report on the following: STEMI and NSTEMI patient demographics; provider and facility characteristics; adverse event rates; AMI performance measures and select quality measures and outcomes; medication dosing errors and risk adjusted metrics; transfer facility therapies and reperfusion strategies; compliance with ACC/AHA clinical guideline recommendations.

The majority of the required data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions can manually report using a free web-based tool or automate the reporting by using certified software developed by third-party vendors. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

There are a number of methods used to educate and provide general support to registry participants. This includes the following:

- Registry Site Manager Calls are available for all NCDR participants. RSM calls are provided as a source of communication between NCDR and participants to provide a live chat Q and A session on a continuous basis.
- New User Calls are available for NCDR participants, and are intended for assisting new users with their questions.
- NCDR Annual Conference
- The NCDR Annual Conference is a well-attended and energetic two-day program at which participants from across the country come together to hear about new NCDR and registry-specific updates. During informative general sessions, attendees can learn about topics such as transcatheter therapies, the NCDR dashboard, risk models, data quality and validation, and value-based purchasing. Attendees also receive registry updates and participate in advanced case studies covering such topics as Appropriate Use Criteria and outcomes report interpretation.
- Release notes (for outcomes reports)
- Clinical Support

The NCDR Product Support and Clinical Quality Consultant Teams are available to assist participating sites with questions Monday through Friday, 9:00 a.m. - 5:00 p.m. ET.

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

Describe how feedback was obtained.

Feedback is typically obtained through monthly registry site manager monthly calls, ad hoc phone calls tracked with salesforce software, and during registry –specific break-out sessions at the NCDR's annual meeting. Registry Steering Committee members may also provide feedback during regularly scheduled calls.

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

Users for the CathPCI registry reported that some sites have expressed difficulty with identifying certain data elements due to how data is being currently captured. For example, communication of patient details to CR facilities is sometimes challenging to capture. However there are many facilities that seem to have very good processes in place that are integrated with their EHR/EMR.

Users from the ACTION registry reported challenges related to sites being able to implement a process at their facility to streamline compliance to the measure since it requires a multi-provider approach to complete the process. However, many sites have been able to develop quality improvement initiatives to improve their compliance. (Related Reference: Ades PA, Keteyian SJ, Wright JS, et al. Increasing Cardiac Rehabilitation Participation From 20% to 70%: A Road Map From the Million Hearts Cardiac Rehabilitation Collaborative. Mayo Clin Proc. 2017;92:234-42)

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

No other feedback was received from other users.

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

N/A (Measure was not modified since last endorsement)

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

The performance data from 2015 and 2016 for the ACTION registry shows modest improvements from 2015 to 2016. The mean performance for Q4 of 2015 is 62% and the mean performance of Q4 in 2016 is 62.7%. The IQR for the ACTION registry shows a smaller range from 40.3% in Q4 of 2015 to 36.9% in Q4 of 2016.

The CathPCI registry is consistent with the ACTION registry in showing modest improvements from 2015 to 2016 as well. The mean performance for Q4 of 2015 is 52.8% and the mean performance of Q4 of 2016 is 53.8%. The IQR for the CathPCI registry also shows a smaller range from 74.7% in Q4 of 2015 to 67% in Q4 of 2016.

We believe that continued expanded implementation of the measure will lead to greater awareness and accountability among providers and accelerate improvements in referral (and enrollment) rates.

4b2. Unintended Consequences

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

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

No unintended negative consequences have been identified via our testing projects nor have any been reported to us by users of the measure.

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

For CathPCI: An unexpected/good outcome of this measure is that many facilities reexamined their processes to be in compliance with CR parameters and found that their process needed improvement.

For ACTION: An unexpected benefit is the improved patient compliance and commitment for other cardiac care measures which has a positive impact the long term outcomes of the patient.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

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

5b. Competing Measures

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

OR

Multiple measures are justified.

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

N/A

Appendix

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

Attachment Attachment: 0642F_TFPM_Supplement_2017.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American College of Cardiology

Co.2 Point of Contact: Sana, Gokak, comment@acc.org, 202-375-6596-

Co.3 Measure Developer if different from Measure Steward: American College of Cardiology

Co.4 Point of Contact: Esteban, Perla, eperla@acc.org, 202-375-6499-

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.

Randal J. Thomas, MD, MS, FAHA, FACP, Chair; Marjorie King, MD, FACC, FAACVPR, member; Karen Lui, RN, C, MS, FAACVPR, member; Ileana L. Piña, MD, FACC, member;John Spertus, MD, MPH, FACC, member; Neil Oldridge, PhD, FAACVPR

The expert workgroup reviewed the available guidelines and other evidence, proposed and specified measures, responded to comments during peer review and public comment, continues to advise on additional specification of the measure and updates.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2007

Ad.3 Month and Year of most recent revision: 09, 2010

Ad.4 What is your frequency for review/update of this measure? Approximately every 3 years or as needed if evidence changes or due to feedback from implementation

Ad.5 When is the next scheduled review/update for this measure? 06, 2018

Ad.6 Copyright statement: Copyright 2010, American Association for Cardiovascular and Pulmonary Rehabilitation, American College of Cardiology Foundation and the American Heart Association

Ad.7 Disclaimers: These measures and specifications are provided "as is" without warranty of any

kind. Neither the AACVPR, the ACCF, nor the AHA shall be responsible for any

use of these performance measures.

Limited proprietary coding is contained in the measure specifications (online data

supplement) for convenience. Users of the proprietary code sets should obtain all

necessary licenses from the owners of these code sets. The AACVPR, the ACCF, and

the AHA disclaim all liability for use or accuracy of any Current Procedural

Terminology (CPT™) or other coding contained in the specifications.

CPT™ contained in the online data supplement is ©2009 American Medical

Association.

Ad.8 Additional Information/Comments: None

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: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0643

Measure Title: Cardiac Rehabilitation Patient Referral From an Outpatient Setting

Measure Steward: American College of Cardiology

Brief Description of Measure: Percentage of patients evaluated in an outpatient setting who in the previous 12 months have experienced an acute myocardial infarction or chronic stable angina or who have undergone coronary artery bypass (CABG) surgery, a percutaneous coronary intervention (PCI), cardiac valve surgery (CVS), or cardiac transplantation, who have not already participated in an early outpatient cardiac rehabilitation/secondary prevention program for the qualifying event, and who are referred to an outpatient cardiac rehabilitation/secondary prevention program.

Developer Rationale: 1. Cardiac rehabilitation/secondary prevention programs (CR/SP) improve patient outcomes, including quality of life, function, recurrent myocardial infarction, and mortality.

2. CR/SP is underutilized with geographic variability and decreased participation by patients with economic disadvantages, women and older patients.

3. The CR/SP performance measures were developed for use in systematic quality improvement projects to close this treatment gap.

4. Use of systematic referral processes and tools have been shown to increase CR/SP referral.

5. Enrollment and participation in CR/SP, not referral, have been shown to improve patient outcomes. However, referral is necessary for patients to enroll and participate in CR/SP. The strength of provider referral to CR has been shown to correlate with participation in CR.

6. ACC recognizes previous comments from the NQF with regard to the inclusion of patients with Chronic Stable Angina (CSA) in this measure set. Measure authors have discussed these comments and have agreed to include CSA patients in this measure as many of these patients do improve from a symptomatic and functional perspective with exercise training/cardiac rehab. Also, referring patients with CSA to cardiac rehab is an accepted standard of care and is covered by CMS as well. Furthermore, measure authors are concerned that removing CSA from the measure may inadvertently send an incorrect message that it is not expected that providers refer patients to cardiac rehab.

Numerator Statement: Number of patients in an outpatient clinical practice who have had a qualifying event/diagnosis during the previous 12 months, who have been referred to an outpatient Cardiac Rehabilitation/Secondary Prevention

(CR/SP) program. (Note: The program may include a traditional CR/SP program based on face-to-face interactions and training sessions or may include other options such as home-based approaches. If alternative CR/SP approaches are used, they should be designed to meet appropriate safety standards and deliver effective, evidence-based services.)

Denominator Statement: Number of patients in an outpatient clinical practice who have had a qualifying cardiovascular event in the previous 12 months and who do not meet any of the criteria listed in the denominator exclusion section below, and who have not participated in an outpatient cardiac rehabilitation program since the qualifying event/diagnosis.

Denominator Exclusions: Exceptions criteria require documentation of one or more of the following factors that may prohibit cardiac rehabilitation participation: Medical factors (e.g., patient deemed by provider to have a medically unstable, life-threatening condition). Health care system factors (e.g., no cardiac rehabilitation/secondary prevention (CR/SP) program available within 60 min of travel time from the patient's home).

The only exclusion criterion for this measure is noted below: Patients already referred to CR from another provider/facility and/or was participating in CR prior to encounter with provider at the current office/facility.(1) When the provider discusses CR/SP referral with the patient, if the patient indicates that he/she has already been referred to CR/SP, then that provider would not be expected to make another referral. However, the provider should document that information in the medical record. Exceptions criteria require documentation of one or more of the following factors that may prohibit cardiac rehabilitation participation: Medical factors (e.g., patient deemed by provider to have a medically unstable, life-threatening condition). Health care system factors (e.g., no cardiac rehabilitation/secondary prevention (CR/SP) program available within 60 min of travel time from the patient's home).

The only exclusion criterion for this measure is noted below: Patients already referred to CR from another provider/facility and/or was participating in CR prior to encounter with provider at the current office/facility.(1) When the provider discusses CR/SP referral with the patient, if the patient indicates that he/she has already been referred to CR/SP, then that provider would not be expected to make another referral. However, the provider should document that information in the medical record.

Measure Type: Process

Data Source: Electronic Health Records, Registry Data, Paper Medical Records

Level of Analysis: Clinician : Group/Practice, Clinician : Individual, Integrated Delivery System

IF Endorsement Maintenance – Original Endorsement Date: May 05, 2010 **Most Recent Endorsement Date:** Sep 08, 2014

Staff Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence

matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

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

Summary of prior review in 2014

- The developer provides a diagram of referral to cardiac rehabilitation program and its relation to patient's health outcomes: Lower Mortality/Morbidity, Higher Quality of Life, Risk Factor Modification, Improved Function & Exercise Capacity, Improved Medication Adherence, Reduction in Re-Hospitalization Rates, and Cost Effective Care
- The developer cited systematic reviews of six ACCF/AHA guidelines with grading of the evidence for referral to cardiac rehabilitation for different heart disease/conditions. No QQC is provided for each of the six guidelines, but evidence grades are defined.
- The developer cited a Cochrane systematic reviews from 2009 with QQC provided and no evidence of publication bias for total mortality, CV mortality, CABG or PTCA. There was evidence of small study bias for total MI. Benefits are reported as risk ratios (95% CI) that compared participation in CR versus usual care based on meta-analyses from the Cochrane Systematic Review. Those with p value < 0.05 are:
    - For CR versus Usual Care, Follow-up > 12 months- Total Mortality (risk ratio=0.87) and CV Mortality (risk ratio=0.74)
    - For Hospital Admissions versus Usual Care, Follow-up 6-12 months with a risk ratio of 0.69

Changes to evidence from last review

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

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

Updates:

- Although the developer attested there are no changes in evidence since it was endorsed in 2014, a new systematic review guideline and a new study from the Cochrane systematic review was added to the submission listed below:
    1. An AHA/ACC guideline for the management of patients with non-ST-elevation (NSTE) acute coronary syndromes (ACS). This guideline recommends all eligible patients with NSTE-ACS should be referred to a comprehensive cardiovascular rehabilitation program either before hospital discharge or during the first outpatient visit.
    2. A new study conducted from the Cochrane systematic review supports the conclusions of the prior review in 2014 that, compared with no exercise control, exercise-based CR reduces the risk of cardiovascular mortality but not total mortality.

Exception to evidence: N/A

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

Questions for the Committee:

- *The evidence provided by the developer is updated and directionally the same compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance from the Evidence Algorithm

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

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

RATIONALE: N/A

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided 2015-2016 performance scores from the ACC PINNACLE registry by mean and decile. The mean result for the two years are below:
    1. For 2015, 5.51% from 4,954 providers and 27,0448 patients
    2. For 2016, 5.42% from 2,752 providers and 21,6773 patients

Disparities

- The developer provides 2015-2016 disparities data by mean results and decile. The data were stratified by gender, age, insurance status, and race as mean results and decile. The mean results are shown below:

Year 2015

Label	Number of Providers	Number of Patients	Mean
Male	4215	172589	5.56%
Female	4106	91280	5.41%
Age: <60	3783	51043	5.71%
Age: 60-<70	3954	74374	5.56%
Age: 70 - <80	3882	83157	5.42%
Age: >=80	3619	55701	5.35%
Insurance: None	755	4084	9.73%
Insurance: Private	3491	113321	6.57%
Insurance: Medicaid	1856	13137	4.62%
Insurance: Medicare	3613	108596	5.97%
Insurance: Other	1439	9593	6.57%
Race: White	3816	166835	5.70%
Race: Black	2072	11591	3.85%
Race: Other	1	3	0.00%

Year 2016

Label	Number of Providers	Number of Patients	Mean
Male	2302	136349	1.23%
Female	2253	72778	1.05%
Age: <60	2118	41225	1.09%
Age: 60-<70	2179	59007	1.18%

Label	Number of Providers	Number of Patients	Mean
Age: 70 - <80	2173	64731	1.18%
Age: >=80	2074	44993	1.08%
Insurance: None	603	5486	1.56%
Insurance: Private	2041	106644	1.19%
Insurance: Medicaid	1134	13066	0.87%
Insurance: Medicare	2077	104772	1.23%
Insurance: Other	1066	6712	0.00%
Race: White	2129	128433	0.52%
Race: Black	1323	10086	0.87%
Race: Other	0	0	0.00%

- Developer noted findings from literature review on disparities that included:
    1. According to a CR referral study in the MI patient population, referral rates improved across gender and racial/ethnic groups, but continued to remain high in males and whites.
    2. Gaps in delivery of cardiac rehabilitation have been documented in the published literature, particularly prevalent in women, the elderly, and in racial/ethnic minority groups.

Questions for the Committee:

- *Does the data demonstrate a quality problem related to patients that do not receive a referral to cardiac rehabilitation from an outpatient setting after a cardiac event?*
- *Is a national performance measure still warranted?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

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

RATIONALE: N/A

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

1a. Evidence:

- Update to evidence supports prior evidence.
- This measure was endorsed in 2010 and 2014. Evidence remains and doesn't need discussion.
- This is a process measure that has intuitive validation but represents a variety of cardiac conditions and a non-uniform cardiac rehabilitation program. Still it has sufficient belief in the outcomes of this process measure that it has been incorporated into CMS's MIPS program
- There is good direct evidence for this measure to improve outcomes through CR referral. I would argue that CSA patients may or may not benefit from CR depending on their history and clinical course and think this should be removed unless some additional qualifications are made.
- Moderate level of evidence

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

- Gaps and disparities exist. Overall CR participation is low, especially for minorities.
- If I am reading the table correctly, the performance gap is huge.

- The ACC/AHA paper describing this measure as published in Circulation makes mention of "gap" only once and that was at a high level. While there is a recognition of the variation in referral rates, the true ideal referral rate is not a standard measure.
- There is a significant performance gap which would justify this metric as offering potential clinical improvement for some outpatient facilities. Possible disparities for this measure are broader than the data and demographic information being collected. Here transportation (which might not just be an economic barrier- could be distance or ability to drive/have driver).
- Moderate performance gap

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing
2b. Validity: Testing; Exclusions**;** Meaningful Differences**;** Comparability**;** Missing Data

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: NQF Staff

Evaluation of Reliability and Validity

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Evaluation of Scientific Acceptability

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.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 0643

Measure Title: Cardiac Rehabilitation Patient Referral From an Outpatient Setting

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
   NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.
   TIPS: Consider the following: 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?
   ☒Yes (go to Question #2)
   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.
2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?
   TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)
   ☒Yes (go to Question #4)
   ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

3. Was **empirical VALIDITY testing** of patient-level data conducted?
   ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)

4. Was reliability testing conducted with computed performance measure scores for each measured entity?
   TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data
   ☒Yes (go to Question #5)
   ☐No (go to Question #8)
5. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? *NOTE: If multiple methods used, at least one must be appropriate.*
   TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.
   ☒Yes (go to Question #6)
   ☐No (please explain below then go to Question #8)
6. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?
   TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?
   Do the results demonstrate sufficient reliability so that differences in performance can be identified?
   ☐High (go to Question #8)
   ☒Moderate (go to Question #8)
   ☐Low (please explain below then go to Question #7)
7. Was other reliability testing reported?
   ☒Yes (go to Question #8)
   ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)
8. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?
   TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)
   ☒Yes (go to Question #9)
   ☐No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?
   TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements
   Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)
   ☒Yes (go to Question #10)
   ☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)
10. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?
    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?
    ☒Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)
    ☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)
    ☐Insufficient (go to Question #11)

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

1. Were all potential threats to validity that are relevant to the measure empirically assessed?
   TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

   ☒Yes (go to Question #2)

   ☐No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]
2. Analysis of potential threats to validity: Any concerns with measure exclusions?
   TIPS: Consider the following: 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)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

   ☒Yes (please explain below then go to Question #3)

   ☐No (go to Question #3)

   ☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

   The developer highlighted that the documentation and assessment of exclusions help mitigate potential bias in reporting (i.e., excluding patients who are actually eligible for cardiac rehabilitation referral, in order to improve performance scores). Based on the results of the Cardiac Rehabilitation Referral Reliability Testing (CR3) project, the time and effort to assess exclusions does not appear to add significant burden.
3. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

   ☒Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

   a. Is a conceptual rationale for social risk factors included? ☐Yes ☐No

   b. Are social risk factors included in risk model? ☐Yes ☐No

   c. Any concerns regarding the risk-adjustment approach?

   *TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? 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, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you*

agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?

☐Yes (please explain below then go to Question #4)

☐No (go to Question #4)

4. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

   ☐Yes (please explain below then go to Question #5)

   ☒No (go to Question #5)

5. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

   ☐Yes (please explain below then go to Question #6)

   ☐No (go to Question #6)

   ☒Not applicable (go to Question #6)

6. Analysis of potential threats to validity: Any concerns regarding missing data?

   ☐Yes (please explain below then go to Question #7)

   ☒No (go to Question #7)

ASSESSMENT OF MEASURE TESTING

7. Was empirical validity testing conducted using the measure as specified and appropriate statistical test?
   Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

   ☐Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

   ☒No (please explain below then go to Question #8)

8. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?
   TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

   ☒Yes (go to Question #9)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

9. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

   ☐Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☒ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

10. Was validity testing conducted with computed performance measure scores for each measured entity?

TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

☐Yes (go to Question #11)

☐No (please explain below and go to Question #13)

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

    TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

    ☐Yes (go to Question #12)

    ☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

12. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #14)

    ☐Moderate (go to Question #14)

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

13. Was other validity testing reported?

    ☐Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

14. Was validity testing conducted with patient-level data elements?

    TIPS: Prior validity studies of the same data elements may be submitted

    ☐Yes (go to Question #15)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

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

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #16)

    ☐No (please explain below and rate Question #16 as INSUFFICIENT)

16. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

    ☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

    ☐Insufficient (go to Question #17)

17. OVERALL VALIDITY RATING

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

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

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

2a. Reliability-Specifications:

- Reliability is moderate. No concerns.
- The variety of diagnoses (i.e. types of heart disease) all have different physiology so make measurement difficult. This is compounded by lack of standardized cardiac rehab. If the elements of cardiac rehab for such patients were more standard (as in the diabetes prevention project) then this would be a more stable measure.
- Although the qualifying descriptors are fairly uniform- the CSA patients may see variability that is even greater in the outpatient (non-acute) setting. In other words they should not be included under the same descriptor.
- Low reliability

2a2. Reliability-Testing:

- Signal to noise ratio is quite high . There was moderate inter-rater reliability of data elements.
- No concerns
- Only as above
- Low reliability

2b2. Validity testing & 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data):

- Face validity is substantial.
- Validity is face. No testing was done, but testing is probably not necessary.
- The greater threat is the lack of consistency of type of heart disease and the characteristics of the rehab program.

- Outpatient physicians will by the developers own admission, not consistently report in the PINNACLE registry for this measure. Due to this it decreases the validity of the measure as outcomes may be unchanged just based on variability in data capture.
- Low validity

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

- Not risk adjusted
- For this measure the patients are "high risk" but the data is not risk adjusted from my review.
- The risk adjustment will include social risk factors but likely not in depth enough based on the description. With the length of reporting and the complex issues that could affect referral (transportation independent of economic or demographic variability) this measure may not adequately be risk adjusted.
- Data?

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent
3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Some of the data elements are captured in electronic clinical data, but the developer mentioned that ACC is currently developing a common data dictionary mapped to coded terminology standards that may improve interoperability with EHRs and potentially create eMeasures. However, the developer states that the data are abstracted from a record by a third party other than the individual obtaining the original information.
- Developers noted three challenges to the feasibility of the measures and provided the ongoing solutions below:
    1. For the ACC PINNACLE registry, project managers with clinical backgrounds provide significant support to local practices that collect data for PINNACLE. Unfortunately even with this support, the developer emphasized that clinicians still do not document this measure even if a patient has been referred to CR. The developer hopes that CMS will incentivize clinicians to ensure that both eligible patients are being referred to a rehabilitation facility and to ensure the measure is documented correctly.
    2. For the ACC Cardiology Practice Improvement Pathway, developers identified the difficulty in capturing the difference between individuals who already participate in CR versus individuals who were referred to CR because of the long measurement window.
    3. For the CR3 project, the developer highly suggested the use of a stricter definition of referral that includes all three components. The developer hypothesized this refined definition will increase the enrollment in CR/SP. However, current practices and existing registries required the first component only. The three components are:
        a) Documentation that patient was referred.
        b) Communication (electronic/written) that referral information was given to patient.
        c) Communication (electronic/written) that the receiving CR site was given patient's referral information.
- No fees, licensing, or other requirements mentioned.

Questions for the Committee:

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

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

RATIONALE: N/A

Committee Pre-evaluation Comments: Criteria 3: Feasibility

3. Feasibility:

- Documentation may be an issue.
- The measure reportedly suffers from under documentation.
- This is being used in the MIPS program so is feasible
- Primarily, the developer has indicated difficulty in getting providers to document or include CR referral in the PINNACLE registry. Also, the length of time for inclusion and possible confusion about when the patient was actually referred create some feasibility problems.
- Not feasible...low feasibility

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- This measure is in use for public reporting, Physician Compare, payment program, and quality improvement. The quality improvement program, specifically PINNACLE Registry is in use for benchmarking or specific to an organization. The developer noted that the Merit-based Incentive Payment System (MIPS), a payment program, is part of the quality payment program (QPP).
- The developer indicate planned use is public reporting and hope to expand the use of this measure in other payment programs (e.g., accountable care organizations, Medicare Advantage insurance plans, other health plans on the insurance marketplace).
- The developer noted that physicians who report this measure for MIPS/QPP have also agreed to report it on Physician Compare.

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

Feedback on the measure by those being measured or others

- Performance results are provided monthly to PINNACLE participants through the PINNACLE Physician Dashboard.
- This measure have not been modified since last endorsement on September 8, 2014.

Additional Feedback:

- Based on the NQF's Cardiovascular report in 2014, some of the Committee members had concerned that providers may be penalized by both this measures and measure #0642 (companion measure), specifically if a patient is referred to cardiac rehabilitation prior to discharge from an inpatient admission but has not enrolled prior to the outpatient visit with the same provider. In conclusion, the Committee voted to continued endorsement of this measure.

***Questions for the Committee**:*

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

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

RATIONALE: N/A

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

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

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

Improvement results

- The developer noted that there appears to be no meaningful improvement from 2015-2016 (5.5% versus 5.6%). Although there appears no improvement, a literature review study conducted on the trends in referral to cardiac rehabilitation after myocardial infarction showed statistical significance in the increase referral rates from 2007-2012. Another study conducted in various states by the Centers for Disease Control and Prevention, found that patients who use outpatient cardiac rehabilitation after a heart attack were 53% less likely to experience cardiac-related mortality compared to those who did not use cardiac rehabilitation. Based on these literature reviews, the developer believed the expanded use of this measure "will lead to greater awareness and accountability among providers and accelerate improvements in referral (and enrollment) rates".

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

Unexpected findings (positive or negative) during implementation

- No unexpected positive and negative findings are reported.

Potential harms

- The developer did not report any unintended consequences.

Additional Feedback:

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

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

RATIONALE: N/A

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

4a. Use:

- The measure is both publicly reported and used in an accountability program.
- This is a process measure being intended to create more referrals to some form of cardiac rehab, as such it is a measure that is more intended to stimulate action than a true measure that can be used for discrimination between providers at a granular level.
- The participants are receiving the performance results but the opportunity (at least systematically) is not evident. It is publicly reported and COMPARE and MIPS are being used for accountability.
- Low usability and use.

4b. Usability:

- I am confused by the performance tables and would like clarification from the developers.
- I have not seen yet studies demonstrating that increase in the referral rate as an outpatient results in lower morbidity and mortality (as opposed to review of the results of patients already engaged in rehab

Mampuya WM. Cardiac rehabilitation past, present and future: an overview. Cardiovascular Diagnosis and Therapy. 2012;2(1):38-49.

- The main unintended consequences might be financial penalties for outpatient physicians just due to their lack of data input despite appropriate referral. Also, referral for CSA patients may be greater than needed without a proven benefit.
- Low

Criterion 5: Related and Competing Measures

Related or competing measures

- There may be related measures and are listed below:
  - 0071: Persistence of Beta-Blocker Treatment After a Heart Attack
  - 0090: Emergency Medicine: 12-Lead Electrocardiogram (ECG) Performed for Non-Traumatic Chest Pain
  - 0137: ACEI or ARB for left ventricular systolic dysfunction- Acute Myocardial Infarction (AMI) Patients
  - 0142: Aspirin prescribed at discharge for AMI
  - 0230: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older.
  - 0290: Median Time to Transfer to Another Facility for Acute Coronary Intervention
  - 0642: Cardiac Rehabilitation Patient Referral From an Inpatient Setting
  - 0730: Acute Myocardial Infarction (AMI) Mortality Rate
  - 0964 Therapy with Aspirin, P2Y12 Inhibitor, and Statin at Discharge Following PCI in Eligible Patients
  - 2377: Defect Free Care for AMI
  - 2379: Adherence to Antiplatelet Therapy after Stent Implantation
  - 2452 PCI: Post-Procedural Optimal Medical Therapy [clinician]
  - 2473: Hospital 30-Day Risk-Standardized Acute Myocardial Infarction (AMI) Mortality eMeasure
- There are no competing measures.

Harmonization

- Committee recommendations for combining or harmonizing measures may be solicited.

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

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 10, 2018

No comments have been submitted as of this date.

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

0643_NQF_evidence_attachment_Sep2017.pdf

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission? Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1a Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0643

Measure Title: **Cardiac Rehabilitation Referral from an Outpatient Setting**

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: Click here to enter a date

Instructions

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

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

1a. Evidence to Support the Measure Focus

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

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.

- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

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

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

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

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

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

Outcome

☐ Outcome:

☐Patient-reported outcome (PRO):

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

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

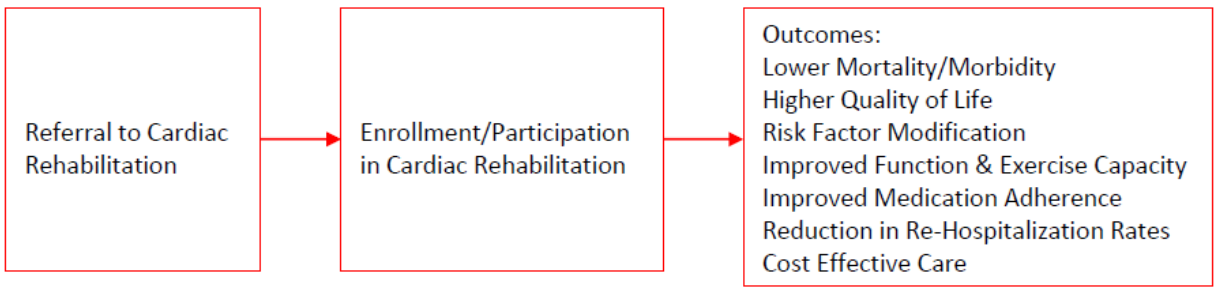
☒ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.



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

**RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or1a.4) **

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Coronary Artery Bypass Surgery (CABG): Hillis LD, et. Al. 2011 ACCF/AHA guideline for coronary artery bypass graft surgery: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2011;124:e652– e735. http://circ.ahajournals.org/content/124/23/e652.full.pdf+html
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Pages e683-684: 4.9. Cardiac Rehabilitation: Cardiac rehabilitation is recommended for all eligible patients after CABG.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	Cardiac rehabilitation is recommended for all eligible patients after CABG. Recommendation Class I; Level of Evidence: A
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA

Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Percutaneous Coronary Intervention (PCI): Levine GN, 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. *J Am Coll Cardiol* 2011;58:e44–122. http://content.onlinejacc.org/article.aspx?articleid=1147816
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Page e89: 6.4.3. Cardiac Rehabilitation: Recommendation Medically supervised exercise programs (cardiac rehabilitation) should be recommended to patients after PCI, particularly for moderate- to high-risk patients for whom supervised exercise training is warranted.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	Medically supervised exercise programs (cardiac rehabilitation) should be recommended to patients after PCI, particularly for moderate- to high-risk patients for whom supervised exercise training is warranted. Recommendation Class I; Level of Evidence: A
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Coronary Artery Disease (CAD): Smith SC Jr., et. Al. 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
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Page 2436 1. All eligible patients with ACS or whose status is immediately post coronary artery bypass surgery or post-PCI should be referred to a comprehensive outpatient cardiovascular rehabilitation program either prior to hospital discharge or during the first follow-up office visit. 2. All eligible outpatients with the diagnosis of ACS, coronary artery bypass surgery or PCI, chronic angina… within the past year should be referred to a comprehensive outpatient cardiovascular rehabilitation program. 3. A home-based cardiac rehabilitation program can be substituted for a supervised, center-based program for low-risk patients.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA

Grade assigned to the **recommendation** with definition of the grade	All eligible patients with ACS or whose status is immediately post coronary artery bypass surgery or post-PCI should be referred to a comprehensive outpatient cardiovascular rehabilitation program either prior to hospital discharge or during the first follow-up office visit. Recommendation Class I; Level of Evidence: A All eligible outpatients with the diagnosis of ACS, coronary artery bypass surgery or PCI, chronic angina, and/or peripheral artery disease within the past year should be referred to a comprehensive outpatient cardiovascular rehabilitation program. PCI: Recommendation Class I; Level of Evidence: A Chronic Angina: Recommendation Class I; Level of Evidence: B A home-based cardiac rehabilitation program can be substituted for a supervised, center-based program for low-risk patients. Recommendation Class I; Level of Evidence: A
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	ST Elevation Myocardial Infarction (STEMI): O'Gara PT, 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:e78 –140, doi:10.1016/j.jacc.2012.11.019. http://content.onlinejacc.org/article.aspx?articleid=1486115
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	e114-116: 11.1. Post hospitalization Plan of Care: Recommendations 1. Post hospital systems of care designed to prevent hospital readmissions should be used to facilitate the transition to effective, coordinated outpatient care for all patients with STEMI. 2. Exercise-based cardiac rehabilitation/secondary prevention programs are recommended for patients with STEMI. 3. A clear, detailed, and evidence-based plan of care that promotes medication adherence, timely follow- up with the healthcare team, appropriate dietary and physical activities, and compliance with interventions for secondary prevention should be provided to patients with STEMI
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	Post-hospital systems of care designed to prevent hospital readmissions should be used to facilitate the transition to effective, coordinated outpatient care for all patients with STEMI. Recommendation Class I; Level of Evidence: B Exercise-based cardiac rehabilitation/secondary prevention programs are recommended for patients with STEMI. Recommendation Class I; Level of Evidence: B A clear, detailed, and evidence-based plan of care that promotes medication adherence, timely follow-up with the healthcare team, appropriate dietary and physical activities, and compliance with interventions for secondary prevention should be provided to patients with STEMI. Recommendation Class I; Level of Evidence: C

Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Amsterdam EA, Wenger NK, Brindis RG, 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. J Am Coll Cardiol. 2014;64:e139-228/ http://www.onlinejacc.org/content/64/24/e139.
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	e139-228 Cardiac Rehabilitation and Physical Activity: Recommendation: All eligible patients with NSTE-ACS should be referred to a comprehensive cardiovascular rehabilitation program either before hospital discharge or during the first outpatient visit. (Class I, Level of Evidence: B)
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	All eligible patients with NSTE-ACS should be referred to a comprehensive cardiovascular rehabilitation program either before hospital discharge or during the first outpatient visit. (Class I, Level of Evidence: B)
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Stable Ischemic Heart Disease (Stable IHD): Fihn SD, 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, 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://circ.ahajournals.org/content/early/2012/11/19/CIR.0b013e318277d6a0.full.pdf
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Pages e91-92: 4.4.1.4. PHYSICAL ACTIVITY 1. For all patients, the clinician should encourage 30 to 60 minutes of moderate-intensity aerobic activity, such as brisk walking, at least 5 days and preferably 7 days per week, supplemented by an increase in daily lifestyle activities (e.g., walking breaks at work, gardening, household work) to improve cardiorespiratory fitness and move patients out of the least-fit, least-active, high-risk cohort (bottom 20%). 3. Medically supervised programs (cardiac rehabilitation) and physician-directed, home-based programs are recommended for at risk patients at first diagnosis.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NA
Provide all other grades and definitions from the evidence grading system	NA
Grade assigned to the **recommendation** with definition of the grade	For all patients, the clinician should encourage 30 to 60 minutes of moderate-intensity aerobic activity, such as brisk walking, at least 5 days and preferably 7 days per week, supplemented by an increase in daily lifestyle activities (e.g., walking breaks at work, gardening, household work) to improve cardiorespiratory fitness and move patients out of the least-fit, least-active, high-risk cohort (bottom 20%). Recommendation Class I; Level of Evidence: B Medically supervised programs (cardiac rehabilitation) and physician-directed, home-based programs are recommended for at risk patients at first diagnosis. Recommendation Class I; Level of Evidence: A
Provide all other grades and definitions from the recommendation grading system	See Table 1 below.

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	NA
Estimates of benefit and consistency across studies	NA
What harms were identified?	NA
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	NA

Table 1 (applies to all Provide all other grades and definitions from the evidence grading system)

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 a 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 considered may/might be reasonable usefulness/effectiveness is unknown/unclear/ uncertain or not well established	COR III: No Benefit / COR III: Harm is not recommended / potentially harmful is not indicated / causes harm should not be performed/ administered/ other / associated with excess morbidity/ mortality is not useful/ beneficial/ effective / 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.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Heran BS,et al. Exercise-based cardiac rehabilitation for coronary heart disease. *Cochrane Database of Systematic Reviews* 2011, Issue 7. Art. No.: CD001800. DOI: 10.1002/14651858.CD001800.pub2. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001800.pub2/pdf **Citation and URL for methodology for evidence review and grading**: The systematic review identified quality of evidence based on risk of bias. System for determining risk of bias was explained in Chapter 8 of Cochrane Handbook for Systematic Reviews for Interventions, 5.0.2, updated September 2009 http://www.mrc-bsu.cam.ac.uk/cochrane/handbook502 .
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**The information in the following questions in this section is based on the Cochrane Systematic Review cited in the source of the systematic review.** **Intervention/Service**: The effectiveness of exercise-based cardiac rehabilitation on mortality, morbidity and health-related quality of life of patients with CHD is addressed. Exercise-based cardiac rehabilitation is defined as a supervised or unsupervised inpatient, outpatient, or community- or home-based intervention including some form of exercise training that is applied to a cardiac patient population. The intervention could be exercise training alone or exercise training in addition to psychosocial and/or educational interventions (i.e. “comprehensive cardiac rehabilitation”). Usual care could include standard medical care, such as drug therapy, but did not receive any form of structured exercise training or advice. **Outcomes**: Total mortality; Total MI; Total revascularizations; Total hospitalizations; Health-related quality of life; Costs and cost-effectiveness

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	An overall grade of methodological quality was not assigned. In the systematic review, individual study quality was graded on a scale for risk of bias. **Allocation:** Nearly all the trial publications simply reported that the trial was "randomized" but did not provide any details. A total of 8/47 (17%) studies reported details of appropriate generation of the random sequence and 7/47 (15%) studies reported appropriate concealment of allocation. **Blinding:** For exercise-based cardiac rehabilitation trials, it is not possible to blind patients and clinicians to the intervention. For the large majority of studies, insufficient information was provided to evaluate the blinding of assessors; only 4 of 47 (9%) reported that outcome assessors were blind to group allocation. **Incomplete outcome data:** Losses to follow-up and drop out were relatively high, ranging from 21% to 48% in 12 trials. Follow-up of 80% or more was achieved in 33/47 (70%) studies. Furthermore, reasons for loss to follow and dropout were often not reported. Two trials did not report information on losses to follow-up. Several trials have excluded significant numbers of patients post-randomization, and thus in an intention to treat analysis, these then have been regarded as dropouts. **Selective reporting:** A number of the included studies were not designed to assess treatment group differences in morbidity and mortality (as these were not the primary outcomes of these trials) and, therefore, may not have fully reported all clinical events that occurred during the follow-up period. All studies collecting validated health-related quality of life outcomes fully reported these outcomes. **Quality of the evidence:** We found no evidence of publication bias for total mortality, CV mortality, CABG or PTCA. There was evidence of small study bias for total MI.

Provide all other grades and definitions from the evidence grading system	Two reviewers (BSH, JMHC) independently assessed the risk of bias in included studies using the Cochrane Collaboration's recommended tool, which is a domain-based critical evaluation of the following domains: sequence generation; allocation concealment; blinding of outcome assessment; incomplete outcome data; and selective outcome reporting. Only author's recommendations were provided: In medium to longer term (i.e. 12 or more months follow-up) exercise-based cardiac rehabilitation is effective in reducing overall and cardiovascular mortality and appears to reduce the risk of hospital admissions in the shorter-term (< 12 months follow-up) in patients with CHD. The available evidence does not demonstrate a reduction in the risk of total MI, CABG or PTCA with exercise based cardiac rehabilitation as compared to usual care at any duration of follow-up. Exercise-based cardiac rehabilitation should be recommended for patients similar to those included in the randomized controlled trials
Grade assigned to the **recommendation** with definition of the grade	NA
Provide all other grades and definitions from the recommendation grading system	NA

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity-Seventeen studies (26 publications) met the inclusion criteria and had extractable data to assess the effects of exercise-based cardiac rehabilitation, compared with usual care, on mortality and morbidity in patients with CHD. These were added to the 30 studies (55 publications) from the original Cochrane review for a total of 47 studies (81 publications). Randomized controlled trials (RCTs) of exercise-based cardiac rehabilitation versus usual care with a follow-up period of at least six months. A total of 47 RCTs, with 10,794 patients. Quality-Trial sample sizes varied widely from 28 to 2304, with a median intervention duration of three (range 0.25 to 30) months and a follow-up of 24 (range six to 120) months. Nearly all the trial publications simply reported that the trial was "randomized" but did not provide any details. For exercise-based cardiac rehabilitation trials, it is not possible to blind patients and clinicians to the intervention. For the large majority of studies, insufficient information was provided to evaluate the blinding of assessors; only 4 of 47 (9%) reported that outcome assessors were blind to group allocation. Losses to follow-up and drop out were relatively high, ranging from 21% to 48% in 12 trials. Follow-up of 80% or more was achieved in 33/47 (70%) studies. Based on funnel plot analysis, no publication bias was found for all cause mortality, cardiovascular mortality, CABG and PTCA. However, there appears to an absence of negative-result trials of small to medium size for MI which was statistically significant (P = 0.019).

Estimates of benefit and consistency across studies	Predictors of all-cause mortality, cardiovascular mortality, recurrent MI, and revascularisation (CABG and PTCA) were examined using univariate meta-regression. …a reduction in both total and cardiac mortality was observed in CHD patients randomized to exercise-based rehabilitation. However, this updated review shows that this mortality benefit is limited to studies with a follow-up of greater than 12months. We also found that with exercise the rate of hospital readmissions may be reduced in studies up to 12 months follow-up (based on 4 trials with 54/254 versus 73/225 events), but not in longer term follow-up. There was no difference between exercise-based cardiac rehabilitation and usual care groups in the risk of recurrent myocardial infarction or revascularization at any duration of follow-up. The following are risk ratios (95% CI); (p) comparing participation in CR versus usual care based on meta-analyses from the Cochrane Systematic Review. **Total Mortality**; CR vs Usual Care, Follow-up 6-12 months: 0.82 [0.67, 1.01]; (p = 0.061) **Total Mortality**; CR vs Usual Care, Follow-up >12 months: 0.87 [0.75, 0.99]; (p = 0.041); **CV Mortality**; CR vs Usual Care, Follow-up 6-12 months: 0.93 [0.71, 1.21]; (p=0.590) **CV Mortality**; CR vs Usual Care, Follow-up >12 months: 0.74 [0.63, 0.87]; (p= 0.00018) **Fatal and/or nonfatal MI** vs Usual Care, Follow –up 6-12 months: 0.92 [0.70, 1.22]; (p=0.560) **Fatal and/or nonfatal MI** vs Usual Care, Follow – up >12 months: 0.97 [0.82, 1.15]; (p=0.730) **CABG vs Usual Care**, Follow –up 6-12 months: 0.91 [0.67, 1.24]; (p=0.550) **CABG vs Usual Care**, Follow – up >12 months: 0.93 [0.68, 1.27]; (p=0.650) NQF staff enter #/title Version 6.5 05/29/13 9 **PTCA** vs Usual Care, Follow –up 6-12 months: 1.02 [0.69, 1.50]; (p=0.930) **PTCA** vs Usual Care, Follow – up >12 months: 0.89 [0.66, 1.19]; (p=0.420) **Hospital Admissions** vs Usual Care, Follow –up 6-12 months: 0.69 [0.51, 0.93]; (p=0.016) **Hospital Admissions** vs Usual Care, Follow – up >12 months: 0.98 [0.87, 1.11]; (p=0.790) Given both the heterogeneity in outcome measures and methods of reporting findings, a meta-analysis was not undertaken for **health-related quality of life**. In seven out of 10 trials reporting health related quality of life using validated measures there was evidence of a significantly higher level of quality of life with exercise-based cardiac rehabilitation than usual care

What harms were identified?	Although this review did not assess harm, "several studies have documented the safety of exercise based cardiac rehabilitation in patients with documented SIHD. The 2007 AHA Scientific Statement on Exercise and Acute Cardiovascular Events estimates the risk of a major adverse cardiac event (MACE) at 1 in 80,000 patient-hours. This low event rate applies to medically supervised programs that evaluate patients before participation, provide serial surveillance, and are equipped to handle emergencies." Fihn SD, 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, 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://circ.ahajournals.org/content/early/2012/11/19/CIR.0b013e318277d6a0.full.pdf
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Anderson L, Thompson DR, Oldridge N, Zwisler AD, Rees K, Martin N, Taylor RS. Exercise-based cardiac rehabilitation for coronary heart disease. *Cochrane Database of Systematic Reviews* 2016, Issue 1. Art. No.: CD001800. DOI: 10.1002/14651858.CD001800.pub3. From the study: This updated Cochrane review supports the conclusions of the previous version of this review that, compared with no exercise control, exercise-based CR reduces the risk of cardiovascular mortality but not total mortality.

1a.4 OTHER SOURCE OF EVIDENCE

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

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

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or

- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

1. Cardiac rehabilitation/secondary prevention programs (CR/SP) improve patient outcomes, including quality of life, function, recurrent myocardial infarction, and mortality.

2. CR/SP is underutilized with geographic variability and decreased participation by patients with economic disadvantages, women and older patients.

3. The CR/SP performance measures were developed for use in systematic quality improvement projects to close this treatment gap.

4. Use of systematic referral processes and tools have been shown to increase CR/SP referral.

5. Enrollment and participation in CR/SP, not referral, have been shown to improve patient outcomes. However, referral is necessary for patients to enroll and participate in CR/SP. The strength of provider referral to CR has been shown to correlate with participation in CR.

6. ACC recognizes previous comments from the NQF with regard to the inclusion of patients with Chronic Stable Angina (CSA) in this measure set. Measure authors have discussed these comments and have agreed to include CSA patients in this measure as many of these patients do improve from a symptomatic and functional perspective with exercise training/cardiac rehab. Also, referring patients with CSA to cardiac rehab is an accepted standard of care and is covered by CMS as well. Furthermore, measure authors are concerned that removing CSA from the measure may inadvertently send an incorrect message that it is not expected that providers refer patients to cardiac rehab.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

2015 and 2016 Performance Scores

Year	Number of Providers	Number of Patients	Minimum	Mean	Maximum	Lower Quartile	Upper Quartile	Quartile Range	Standard Deviation
2015	4954	270448	0.00%	5.51%	24.18%	1.38%	13.33%	11.95%	17.41%
2016	2752	216773	0.00%	5.42%	26.05%	1.14%	13.50%	12.36%	18.03%

2015 and 2016 Performance Scores by Decile

Year	Decile 10	Decile 20	Decile 30	Decile 40	Median	Decile 60	Decile 70	Decile 80	Decile 90	Maximum
2015	0.00%	0.78%	1.91%	3.66%	5.51%	8.20%	11.24%	16.00%	24.18%	24.18%
2016	0.00%	0.00%	1.91%	3.66%	5.42%	8.00%	11.31%	16.34%	26.05%	26.05%

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.

1. Thomas RJ, Miller NH, Lamendola C, Berra K, Hedbäck B, Durstine JL, Haskell W. National Survey on Gender Differences in Cardiac Rehabilitation Programs. Patient characteristics and enrollment patterns. J Cardiopulm Rehabil. 1996 Nov-Dec;16(6):402-12.

2. Centers for Disease Control and Prevention (CDC). Receipt of outpatient cardiac rehabilitation among heart attack survivors--United States, 2005. MMWR Morb Mortal Wkly Rep. 2008 Feb 1;57(4):89-94.

3. Suaya J, Shepard DS, Normand SL, Ades PA, Prottas J, Stason WB. Use of cardiac rehabilitation by Medicare beneficiaries after myocardial infarction or coronary bypass surgery. Circulation. 2007 Oct 9;116(15):1653-62.

4. Curnier DY, Savage PD, Ades PA. Geographic distribution of cardiac rehabilitation programs in the United States. J Cardiopulm Rehabil. 2005 Mar-Apr;25(2):80-4.

5. Grace SL, Gravely-Witte S, Brual J, Monette G, Suskin N, Higginson L, Alter DA, Stewart DE. Contribution of patient and physician factors to cardiac rehabilitation enrollment: a prospective multilevel study. Eur J Cardiovasc Prev Rehabil. 2008 Oct;15(5):548-56

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

2015 Disparities Data (in percent)

Label	Number of Providers	Number of Patients	Minimum	Maximum	Lower Quartile	Mean	Upper Quartile	Quartile Range	Standard Deviation
Male	4215	172589	0.00	33.33	1.38	5.56	13.33	11.95	17.09
Female	4106	91280	0.00	30.91	1.29	5.41	13.21	11.92	16.24
Age: <60	3783	51043	0.00	31.82	1.18	5.71	14.43	13.25	16.11
Age: 60-<70	3954	74374	0.00	32.43	1.38	5.56	13.55	12.17	16.46
Age: 70 - <80	3882	83157	0.00	32.95	1.38	5.42	12.82	11.44	16.55
Age: >=80	3619	55701	0.00	31.25	1.41	5.35	12.80	11.39	15.53
Insurance: None	755	4084	0.00	36.59	2.56	9.73	22.73	20.17	16.69
Insurance: Private	3491	113321	0.00	33.33	1.64	6.57	14.29	12.65	17.00
Insurance: Medicaid	1856	13137	0.00	29.09	0.67	4.62	12.33	11.66	15.40
Insurance: Medicare	3613	108596	0.00	30.51	1.53	5.97	13.14	11.61	15.97
Insurance: Other	1439	9593	0.00	30.77	1.64	6.57	14.20	12.56	14.94
Race: White	3816	166835	0.00	29.33	1.33	5.70	13.14	11.81	15.93

Label	Number of Providers	Number of Patients	Minimum	Maximum	Lower Quartile	Mean	Upper Quartile	Quartile Range	Standard Deviation
Race: Black	2072	11591	0.00	28.17	0.39	3.85	10.74	10.35	15.09
Race: Other	1	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00

2015 Disparities Data by Decile (in percent)

Label	Decile10	Decile 20	Decile 30	Decile 40	Median	Decile 60	Decile 70	Decile 80	Decile 90	Maximum
Male	0.00	0.00	0.00	0.00	3.30	6.90	11.61	18.67	33.33	33.33
Female	0.00	0.00	0.00	0.00	3.27	6.78	11.43	18.18	30.91	30.91
Age: <60	0.00	0.00	0.00	1.43	4.17	7.69	12.21	18.87	31.82	31.82
Age: 60-<70	0.00	0.00	0.00	1.05	3.90	7.52	11.90	18.75	32.43	32.43
Age: 70 - <80	0.00	0.00	0.00	1.26	4.05	7.69	12.12	18.84	32.95	32.95
Age: >=80	0.00	0.00	0.00	1.82	4.52	8.14	12.50	18.75	31.25	31.25
Insurance: None	0.00	0.00	2.29	4.82	7.69	11.76	16.67	23.76	36.59	36.59
Insurance: Private	0.00	0.00	0.00	1.59	4.65	8.57	12.88	20.00	33.33	33.33
Insurance: Medicaid	0.00	0.00	0.00	2.10	4.82	8.33	12.17	18.02	29.09	29.09
Insurance: Medicare	0.00	0.00	0.00	0.67	3.85	7.69	11.96	18.57	30.51	30.51
Insurance: Other	0.00	0.00	1.90	4.27	7.14	10.08	14.20	19.28	30.77	30.77
Race: White	0.00	0.00	0.00	0.00	2.29	5.71	10.30	17.02	29.33	29.33
Race: Black	0.00	0.00	0.67	2.42	4.78	7.83	11.59	17.17	28.17	28.17
Race: Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

2016 Disparities Data (in percent)

Label	Number of Providers	Number of Patients	Minimum	Maximum	Lower Quartile	Mean	Upper Quartile	Quartile Range	Standard Deviation
Male	2302	136349	0.00	33.33	1.38	1.23	5.47	13.50	17.95
Female	2253	72778	0.00	33.33	1.29	1.05	5.33	13.50	17.51
Age: <60	2118	41225	0.00	35.71	1.18	1.09	5.66	14.62	18.10
Age: 60-<70	2179	59007	0.00	33.33	1.38	1.18	5.42	13.76	17.38
Age: 70 - <80	2173	64731	0.00	34.78	1.38	1.18	5.33	12.96	17.84
Age: >=80	2074	44993	0.00	35.29	1.41	1.08	5.33	12.94	17.51
Insurance: None	603	5486	0.00	42.55	2.56	1.56	5.72	16.52	18.76
Insurance: Private	2041	106644	0.00	35.63	1.64	1.19	6.36	14.47	18.18
Insurance: Medicaid	1134	13066	0.00	32.08	0.67	0.87	5.17	12.50	16.84
Insurance: Medicare	2077	104772	0.00	32.43	1.53	1.23	6.25	13.76	17.80
Insurance: Other	1066	6712	0.00	29.63	1.64	0.00	4.08	12.12	16.09
Race: White	2129	128433	0.00	27.27	1.33	0.52	5.13	11.96	15.70
Race: Black	1323	10086	0.00	33.33	0.39	0.87	4.82	11.54	16.43
Race: Other	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00

2016 Disparities Data by Decile (in percent)

Label	Decile 10	Decile 20	Decile 30	Decile 40	Median	Decile 60	Decile 70	Decile 80	Decile 90	Maximum
Male	0.00	0.00	0.00	0.00	1.59	5.00	9.76	16.67	33.33	33.33
Female	0.00	0.00	0.00	0.00	1.79	5.17	9.90	16.67	33.33	33.33
Age: <60	0.00	0.00	0.00	0.00	2.52	6.00	10.81	17.86	35.71	35.71
Age: 60-<70	0.00	0.00	0.00	0.00	2.15	5.56	10.31	17.05	33.33	33.33
Age: 70 - <80	0.00	0.00	0.00	0.00	2.26	5.72	10.60	17.39	34.78	34.78
Age: >=80	0.00	0.00	0.00	0.00	2.63	6.15	10.81	17.86	35.29	35.29
Insurance: None	0.00	0.00	0.00	1.18	3.77	6.52	12.50	23.53	42.55	42.55
Insurance: Private	0.00	0.00	0.00	0.00	1.89	5.71	10.58	17.31	35.63	35.63

Label	Decile 10	Decile 20	Decile 30	Decile 40	Median	Decile 60	Decile 70	Decile 80	Decile 90	Maximum
Insurance: Medicaid	0.00	0.00	0.00	0.35	3.27	6.57	10.71	16.42	32.08	32.08
Insurance: Medicare	0.00	0.00	0.00	0.00	1.54	5.26	10.00	16.52	32.43	32.43
Insurance: Other	0.00	0.00	0.00	1.06	3.85	7.41	11.29	16.83	29.63	29.63
Race: White	0.00	0.00	0.00	0.00	0.00	3.92	8.11	13.89	27.27	27.27
Race: Black	0.00	0.00	0.00	1.72	4.17	7.14	11.11	17.31	33.33	33.33
Race: Other	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

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

According to a CR referral study in the MI patient population, referral rates improved across gender and racial/ethnic groups, but still remained higher in males and whites.

Beatty AL, Li S, Thomas L, et al. Trends in referral to cardiac rehabilitation after myocardial infarction: data from the National Cardiovascular Data Registry 2007 to 2012. J Am Coll Cardiol. 2014;63:2582-3.

Gaps in delivery of cardiac rehabilitation have been documented in the published literature. That gap is particularly pronounced in women, the elderly, and in racial/ethnic minority groups.

References:

Thomas RJ, Miller NH, Lamendola C, Berra K, Hedbäck B, Durstine JL, Haskell W.

National Survey on Gender Differences in Cardiac Rehabilitation Programs. Patient characteristics and enrollment patterns. J Cardiopulm Rehabil. 1996 Nov-Dec;16(6):402-12.

Suaya JA, Shepard DS, Normand SL, Ades PA, Prottas J, Stason WB. Use of cardiac rehabilitation by Medicare beneficiaries after myocardial infarction or coronary bypass surgery. Circulation. 2007 Oct 9;116(15):1653-62.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Cardiovascular, Cardiovascular : Coronary Artery Disease, Cardiovascular : Coronary Artery Disease (AMI), Cardiovascular : Coronary Artery Disease (PCI), Surgery : Cardiac Surgery

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

Primary Prevention

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

Elderly, Populations at Risk, Populations at Risk : Individuals with multiple chronic conditions

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://content.onlinejacc.org/article.aspx?articleid=1138518

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:** pinn_v1_datadictionaryfullspecifications_1-5.pdf

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

There have been no changes since our submission in 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) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Number of patients in an outpatient clinical practice who have had a qualifying event/diagnosis during the previous 12 months, who have been referred to an outpatient Cardiac Rehabilitation/Secondary Prevention (CR/SP) program. (Note: The program may include a traditional CR/SP program based on face-to-face interactions and training sessions or may include other options such as home-based approaches. If alternative CR/SP approaches are used, they should be designed to meet appropriate safety standards and deliver effective, evidence-based services.)

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

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

All information required to collect/calculate the numerator, including all codes, logic, and definitions):

Qualifying events include all patients who within the past 12 months experienced myocardial infarction (MI), coronary artery bypass graft surgery (CABG), percutaneous coronary intervention (PCI), cardiac valve surgery, heart

transplantation, and/or who have a current diagnosis of chronic stable angina. A referral is defined as an official communication between the healthcare provider and the patient to recommend and carry out a referral order to an outpatient CR program. This includes the provision of all necessary information to the patient that will allow the patient to enroll in an outpatient CR program. This also includes a written or electronic communication between the healthcare provider or healthcare system and the cardiac rehabilitation program that includes the patient's enrollment information for the program. A hospital discharge summary or office note may potentially be formatted to include the necessary patient information to communicate to the CR program (e.g., the patient's cardiovascular history, testing, and treatments). According to standards of practice for cardiac rehabilitation programs, care coordination communications are sent to the referring provider, including any issues regarding treatment changes, adverse treatment responses, or new nonemergency condition (new symptoms, patient care questions, etc.) that need attention by the referring provider. These communications also include a progress report once the patient has completed the program. All communications must maintain an appropriate level of confidentiality as outlined by the 1996 Health Insurance Portability and Accountability Act (HIPAA).

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

Number of patients in an outpatient clinical practice who have had a qualifying cardiovascular event in the previous 12 months and who do not meet any of the criteria listed in the denominator exclusion section below, and who have not participated in an outpatient cardiac rehabilitation program since the qualifying event/diagnosis.

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

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

N/A

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

Exceptions criteria require documentation of one or more of the following factors that may prohibit cardiac rehabilitation participation: Medical factors (e.g., patient deemed by provider to have a medically unstable, life-threatening condition). Health care system factors (e.g., no cardiac rehabilitation/secondary prevention (CR/SP) program available within 60 min of travel time from the patient's home).

The only exclusion criterion for this measure is noted below: Patients already referred to CR from another provider/facility and/or was participating in CR prior to encounter with provider at the current office/facility.(1) When the provider discusses CR/SP referral with the patient, if the patient indicates that he/she has already been referred to CR/SP, then that provider would not be expected to make another referral. However, the provider should document that information in the medical record.Exceptions criteria require documentation of one or more of the following factors that may prohibit cardiac rehabilitation participation: Medical factors (e.g., patient deemed by provider to have a medically unstable, life-threatening condition). Health care system factors (e.g., no cardiac rehabilitation/secondary prevention (CR/SP) program available within 60 min of travel time from the patient's home).

The only exclusion criterion for this measure is noted below: Patients already referred to CR from another provider/facility and/or was participating in CR prior to encounter with provider at the current office/facility.(1) When the provider discusses CR/SP referral with the patient, if the patient indicates that he/she has already been referred to CR/SP, then that provider would not be expected to make another referral. However, the provider should document that information in the medical record.

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

Exceptions:

All eligible patients who can participate in even a low intensity exercise program and who have the cognitive ability to carry out the individualized education and counseling to life-long secondary prevention efforts should be referred to cardiac rehabilitation/secondary prevention programs, because morbidity and mortality benefits extend to nearly all patient populations, regardless of age or co-morbidities. As a result, the exception examples included in the performance measure relate to either the patient's inability to attend an exercise program (due to physical or practical obstacles) or to cognitive deficits which make them unable to actively participate in exercise or to apply secondary prevention recommendations.

Examples, justification, and data collection issues for exceptions for this measure;

1. Medical factors (e.g., patient deemed by provider to have a medically unstable, life-threatening condition): Medically unstable, life-threatening conditions are contraindications to aerobic exercise and require medical efforts to stabilize and reverse those conditions, rather than efforts directed at secondary prevention of cardiovascular disease. Objective criteria for contraindications to exercise training are included in AHA, ACC, and AACVPR statements and guidelines, which are readily available to practicing clinicians and abstractors. After the condition has been stabilized or reversed, then referral to CR/SP is appropriate. Providers document the specific reason for this exception in clinical notes, summaries and problem lists, which can be abstracted.

2. Health care system factors (e.g., no cardiac rehabilitation program available within 60 minutes of travel time from the patient's home): Although some patients may do so, it is not practical to expect a patient to drive for 2 hours 2 or 3 times per week in order to attend a program that lasts for 1 to 2 hours and research has shown that distance to CR/SP is inversely correlated with attendance We chose 60 minutes (assuming average 30 mph driving speed) based on published data showing that the adjusted odds ratio (OR) to attend CR/SP decreased as the distance from patient zip code to nearest CR/SP facility increased, with the greatest decline between 10.2 (6.5-14.9) miles (OR 0.58) to 31.8 (15.0-231.0) miles (OR 0.29). Although alternative delivery models such as those using telemedicine or home care may be developed in future to provide CR/SP, currently there is no reimbursement for these programs. Therefore, it is unreasonable to hold the provider responsible to refer a patient to a program that he/she is highly unlikely to attend. Providers can determine availability of CR/SP programs from on-line or local resources and document this exception in the medical record. Abstractors can verify the exceptions by cross-referencing the patient's address with publicly available lists of CR/SP program locations.

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

Measure was not stratified. Since all patient sub-groups are reported to have low referral rates and low utilization rates for cardiac rehabilitation services, there is no specific requirement to report data on this performance measure in a stratified format. However, medical centers are encouraged to utilize any stratification of their data as they use the performance measure to identify suboptimal processes and also subgroups at particular risk that are under their care. Such stratification could include stratification by gender, ethnicity, and/or age, since these variables have been found to identify subpopulations that are at particular risk for non-referral to CR/SP in some cities and regions.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

ACC PINNACLE Registry Calculation: Practice ID present= YES AND Provider NPI= YES AND Age at start of measurement period is 18 years or older= YES AND Encounter Date is in the reporting date= YES AND Qualifying Event: Myocardial Infarction (within 12 months) =YES OR Qualifying Event: Coronary Artery Bypass Graft (Within 12 months) = YES OR

Qualifying Event: Cardiac Valve Surgery (Within 12 months)= YES OR Qualifying Event: Heart Transplantation =YES OR Qualifying Event: Stable Angina (within 12 months) AND Current Diagnosis= YES OR Qualifying Event: PCI-stent (within 12 months)= YES OR Qualifying Event: PCI- other (non-stent) intervention= YES AND Yes, Patient already participating in rehab= NO AND Cardiac Rehab Referral or Plan for qualifying event/diagnosis in the past 12 months= YES And Referral Plan Documented= YES

AACVPR/ACC/AHA Cardiac Rehabilitation Referral Reliability Testing (CR3): Hospital ID present = YES AND Subject ID = YES AND *Provider NPI = YES AND Age at start of measurement period is 18 years or older = YES AND Qualifying Event: Myocardial Infarction = YES OR Qualifying Event: Coronary Artery Bypass Graft = YES OR Qualifying Event: Cardiac Valve Surgery = YES OR Qualifying Event: Heart Transplantation = YES OR Qualifying Event: Stable Angina = YES OR Qualifying Event: PCI-stent = YES OR Qualifying Event: PCI- other intervention = YES AND Yes, documentation that patient was referred to CR for this event/diagnosis *Since the data for the CR3 Project were processed through the NCDR-PINNACLE Center, NPI was used to help process the data in accordance with the software used at the Center, which requires an NPI on each report. However, since the purpose of the CR3 Project was to assess reliability of the chart abstraction process and not to assess the variability of CR/SP referral by providers, we opted to analyze the CR/SP referral rates by site, and to use the site NPI for data processing purposes only.

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

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

This performance measure is not based on a sample.

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

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

This performance measure is not based on survey or patient-reported data

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

If other, please describe in S.18.

Electronic Health Records, Registry Data

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

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

American College of Cardiology PINNACLE registry and AACVPR/ACC/AHA Cardiac Rehabilitation Testing (CR3) Project.

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

No data collection instrument provided

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

Clinician : Group/Practice, Clinician : Individual, Integrated Delivery System

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

Outpatient Services

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

2.1 For maintenance of endorsement

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

No

2.2 For maintenance of endorsement

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

No

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** 0643
Measure Title: Cardiac Rehabilitation Patient Referral from an Outpatient Setting
Date of Submission: 11/8/2017

Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures**, section **2b3** also must be completed.

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

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

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

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

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

AND

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

Notes

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

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

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

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

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

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

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

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

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

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

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

The following datasets were used: AACVPR/ACCF/AHA Cardiac Rehabilitation Referral and Reliability (CR3) Project and the ACCF/AHA PINNACLE Registry

1.3. What are the dates of the data used in testing? 2009-2012

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

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

AACVPR/ACCF/AHA CR3 Project:

U.S. practices identified from the ACCF, AHA, and AACVPR databases were invited to participate in the Cardiac Rehabilitation Referral and Reliability (CR3) Project. We sought a variety of outpatient practices, based on varied geographical locations, community sizes, and hospital types/sizes. Outpatient practices that met participation criteria were included in the project. Participation criteria included a willingness and ability to: (1) provide a study coordinator and 2 chart abstractors, (2) complete the project within the specified timeline, and (3) obtain local IRB approval to carry out the project in their outpatient practice. Once each hospital completed and submitted their required data, they were sent a small token of appreciation from AACVPR, ACCF, and AHA. A total of 45 outpatient practices expressed an interest in participating in the project, including hospitals from outside the U.S. (Puerto Rico, Romania, and Turkey). 6 outpatient centers (all in the United states and distributed around the country) met all participation criteria and were selected to participate in the project. The sites used a mixture of paper medical records and EHR systems.

ACCF PINNACLE Registry :

Data were analyzed from the ACCF outpatient registry, PINNACLE. The sample populations, for calendar year 2011 and calendar year 2012, include 252,331 patients from 994 practice in 2011 and 298,206 patients from 1022 practices in 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*)

AACVPR/ACCF/AHA CR3 Project:

Descriptive statistics are noted below:

Sex: Male: 65% (n=152 / 234), Female: 35% (n=82/234)

Age: 18-39: 3% (n=7/229), 40-64: 40% (n=91 / 229), 65-79: 45% (n=103 / 229) , 80+: 12% (n=28 / 229)

Race: White: 84% (n=196 / 234), Black: 8% (n=19 / 234), Asian: 0% (n=1 / 234), American Indian: 1% (n=3 / 234), Native Hawaiian/Pacific Islander: 0% (n=1 / 234), Other: 6% (n=14 / 234)

Hispanic Ethnicity: 0% (n=1 / 234)

PINNACLE Registry, 2012:

	Total **n = 252331**
Race	
(1) White	117261 (89.9%)
(2) Black	8758 (6.7%)
(3) Other	4415 (3.4%)
Missing (.)	121897
Insurance	
(0) No insurance	14914 (7.0%)
(1) Private	129907 (61.1%)
(2) Medicare	61289 (28.8%)
(3) Medicaid	3956 (1.9%)
(4) Other	2629 (1.2%)
Missing (.)	39636
Age	
18 to <60	71020 (28.1%)
60 to <70	67696 (26.8%)
70 to <80	65497 (26.0%)
80 to 112	48118 (19.1%)
Sex	
(1) Male	149415 (59.2%)
(2) Female	102812 (40.8%)
Missing (.)	104
BMI	29.7 ± 6.4
Missing	91870
Diabetes	66294 (26.3%)
CAD	247440 (98.1%)
Hypertension	209013 (82.8%)
AFib	59525 (23.6%)
HF	76388 (30.3%)
PAD	89780 (35.6%)
Prior Stroke/TIA	79532 (31.5%)
MI history	125549 (49.8%)

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 described above was used for all aspects of testing.

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

_____

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*)

AACVPR/ACCF/AHA CR3 Project:

The aim of this project was to assess the reliability and feasibility of abstracting the Cardiac Rehabilitation Referral Performance Measure from an outpatient setting. The sites identified to participate in the project were asked to identify one study coordinator and two chart abstractors. 35 patients were identified by the study coordinator at each site from a consecutive sample of patients admitted to their hospital having a qualifying diagnosis for CR, and discharged alive, starting in August 1, 2009. The two abstractors at each site reviewed the same 35 patient records from their hospital twice (once at baseline, and again one week later).

Site coordinators were instructed to include in the 35 patient records, 30 patients who had qualifying diagnoses for CR/SP referral (to capture sensitivity testing) and 5 patients who did not have a qualifying diagnosis for CR/SP referral (to capture specificity testing). The qualifying events are indicated in the measure numerator statement. The non-qualifying events for the purpose of this abstraction project needed to have one or more of the following diagnoses: heart failure, atrial fibrillation, or syncope.

The CR3 Project Workgroup worked with the study coordinators to address reliability, feasibility, and usability properties for the cardiac rehabilitation performance measures. Specifically the workgroup created chart abstraction forms, site coordinator instructions, abstractor instructions, sample IRB protocol, frequently asked questions, and tracking forms to keep track of the intra-rater (1 abstractor reviewing the same patient record two times) and inter-rater process (2 abstractors reviewing the same patient record). The workgroup had a kickoff call with each center's study coordinator to acquaint him/her with the abstraction project. The workgroup communicated weekly with site coordinators to address any questions or comments the sites may have had.

Abstractors reviewed each patient record and completed the CR3 Project form (see supplement).

Definitions used:

Eligible patient: a patient that had a qualifying event/diagnosis during the hospital period under review

Non-eligible patient: a patient that did not have a qualifying event/diagnosis during the hospital period under review

CR/SP referral : documentation in the patient record for the index hospitalization that the patient was being referred to an outpatient cardiac rehabilitation/secondary prevention program

Exception to referral : documentation in the patient record for the index hospitalization that a patient who was eligible for CR/SP referral had a patient, medical, or healthcare system exception that prohibited their participation in CR/SP

Analyses were performed as follows:

1. Intra-rater and inter-rater agreement between patient record reviews

   Eligibility: Was the patient eligible for CR/SP referral?

   CR/SP Referral: Was each eligible patient referred to CR/SP?

   Exceptions: For patients not referred to CR/SP, was/were any exception(s) to CR/SP documented?

2. Percent agreement

   In what percentage of patient record abstractions did the abstractors agree (for both intra-rater and inter-rater agreement)?

3. Kappa statistic

Site specific: Calculated for the 2 abstractors at each site, to compare intra- and inter-abstractor reliability, with regards to his/her assessment of: (1) eligibility for CR/SP referral, (2) referral to CR/SP, and (3) exceptions to CR/SP referral

Pooled estimate: data from all sites were combined to calculate a pooled kappa statistic for intra- and inter-observer reliability for assessing CR/SP eligibility, referral, and exceptions.

By convention, a kappa > .70 is considered acceptable inter-rater reliability.(1) 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.)

It is important to consider both the "percent agreement" and the kappa statistic when assessing the reliability of abstracting this performance measure from patient records, especially for the assessment of "eligibility" and "exceptions". Each method of reliability assessment gives a slightly different view of reliability in this case.

"Percent agreement" is a helpful assessment of reliability of the measure, but given that over 80% of the patients in the study sample were eligible for cardiac rehabilitation, and more than 90% of the patients were free from exceptions to cardiac rehabilitation participation, the percent agreement for the abstractors may have been somewhat inflated, since by chance alone abstractors may have chosen the "right" eligibility or exception status. (To help minimize this, we blinded the abstractors to the actual number/percentage of patients who were eligible for cardiac rehabilitation in their sample. In addition, abstractors were unaware of the range of exceptions that would be expected in their sample.)

The kappa statistic performs best when there is nearly equal chance of study outcomes (for example, equal chance of being eligible or not eligible for cardiac rehabilitation). When there is a high likelihood of one of the two outcomes, as in our study (high likelihood of eligibility), the results of the kappa analyses can sometimes be less accurate and actually underestimate the true reliability the measure due to a phenomenon that is referred to as a "kappa score paradox" in which there is high percent agreement, yet a low kappa score. (Reference: Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. J Clin Epidemiol. 1996 Apr;49(4):431-4.) Indeed, we observed in our site specific analyses that in some centers with very high percent agreement within and between abstractors, the kappa statistics were very low or even zero in some rare cases. With this in mind, the kappa statistic may underestimate the true reliability of the CR measure.

Using both the "percent agreement" and the kappa statistic together provides a robust view of the reliability of the CR performance measure. One ("percent agreement") may slightly overestimate reliability and the other (kappa statistic) may slightly underestimate reliability. The true reliability of the measure most likely lies between the results from the two methods of assessment. Since the "percent agreement" method suggests "almost perfect" reliability and the kappa statistic suggests "substantial" to "almost perfect" reliability, the overall reliability of the CR performance measure appears to be between "substantial" and "almost perfect"

PINNACLE Registry:

Data were used to assess reliability and other performance characteristics for centers participating in the PINNACLE Registry from January 1 2011 until December 31, 2012.

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 five different points: 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*)

AACVPR/ACCF/AHA CR3 Project (pooled data results):

The abstractor and coordinator experiences in chart abstraction prior to participating in the CR3 project varied greatly. The summary data is below.

Less than 1 month 39% (n= 11)

1-6 months 11% (n=3)

6-12 months 7% (n=2)

1-2 yrs 4% (n=1)

2-3 yrs 4% (n=1)

3-4 yrs 11% (n=3)

4-5 yrs___None

More than 5 years 25% (n=7)

ARE PATIENTS ELIGIBLE FOR CARDIAC REHABILITATION?

Percentage deemed eligible for cardiac rehabilitation: 199 / 234 (85%) (mean of all observations)

(Actual percentage of patients who were eligible for cardiac rehabilitation: 200/234 (86%))

Intra-rater reliability (agreement within the same abstractor):

% Agreement: 232 / 232 (100%)

Kappa: 1.00 (-)

Inter-rater reliability (agreement between abstractors):

% Agreement: 218 / 231 (94%)

Kappa: 0.77 (0.65, 0.89)

HAVE PATIENTS BEEN REFERRED TO CARDIAC REHABILITATION?

Percentage referred to cardiac rehabilitation: 111 /185 (60%) (mean of all observations)

Intra-rater reliability:

% Agreement: 172 / 176 (98%)

Kappa: 0.95 (0.90, 0.99)

Inter-rater reliability:

% Agreement: 148 / 172 (86%)

Kappa: 0.70 (0.59, 0.81)

ARE THERE EXCEPTIONS NOTED FOR ELIGIBLE PATIENTS NOT REFERRED TO CARDIAC REHABILITATION?

Percentage with documented exceptions to cardiac rehabilitation: 17 /201 (9%) (mean of all observations)

Intra-rater reliability:

% Agreement: 189 / 196 (96%)

Kappa: 0.76 (0.60, 0.93)

Inter-rater reliability:

% Agreement: 185 / 191 (97%)

Kappa: 0.79 (0.63, 0.95)

PINNACLE, 2012:

Description	Number of Patients	Signal-to-Noise Ratio
Minimum	10	0.990
25th percentile	87	0.995
50th percentile	173	0.998
75th percentile	379	0.998
Average	292	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?*)

The CR3 project demonstrates high to very high reliability of the measure. The PINNACLE data analysis demonstrates excellent reliability when evaluated at the minimum level of quality reporting events and higher reliability at the median number of events (50th%), and at average and greater number of quality events.

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☐ **Empirical validity testing- Will aim to obtain additional empirical validity testing data for future measures as time allows**

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

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

Face and Content Validity

Validity Survey of Experts: Validity of the measure score was systematically assessed as follows: After the measure was fully specified, members of 3 existing committees, one at the ACC, one at AHA and one joint ACC/AHA, with expertise in general cardiology, cardiac rehabilitation, quality improvement, outcomes research, 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

Face and content validity of the measure score was systematically assessed according to responses received from survey respondents.

AACVPR/ACCF/AHA performance measures set: To determine the content/context validity of the measures, a process using a Delphi peer review was utilized. An explicit and standardized process for ACCF/AHA performance measure development was followed, including the following steps: 1. Formation of the Development Committee, 2. Identification of Potential Factors, for Inclusion, 3. Scoring of the Factors/Expert Opinion, 4. Public Comment Period/Peer Review, 5. Further Refinement, 6. Final approval by organizations, 7. Peer Review Publication/Endorsement. Reviewers were asked to provide comments on the document on the basis of the rating form and guide shown on page 1432 at Http://content.onlinejacc.org/cgi/reprint/j.jacc.2007.04.033v1.pdf

Content/context validity of the measures was also established by virtue of the specialized expertise of the Performance Measures Work Group members including the structured discussions that the work group conducted, as well as the rigorous peer review and public comment period that were carried out. For this particular topic those individuals who were involved in identifying and drafting the performance measures were leaders and experts in the field of cardiac rehabilitation as chosen by AACVPR, ACCF, and AHA.

Furthermore, additional face and content validity was demonstrated from the update of the measure in 2010. During the NQF Care Coordination project, the Steering Committee asked AACVPR, ACCF, and AHA to remove patient refusal as an exception. Since that time, all 3 organizations have published an updated document (NQF measures 0642 and 0643) that explicitly notes that patient refusal should not be an allowable exception. In addition, the cardiac rehabilitation referral measures were revised to facilitate the implementation of these two measures by including administrative codes to identify denominator-eligible populations. All changes were approved by the American Association of Cardiovascular and Pulmonary Rehabilitation Board of Directors, the American College of Cardiology Foundation Board of Trustees, and by the American Heart Association Science Advisory and Coordinating Committee. The performance measure set was also reviewed via AHA and ACC processes as well as by the AACVPR Document Oversight Committee.

AACVPR/ACCF/AHA CR3 Project: Through the NQF endorsement process, the cardiac rehabilitation referral performance measures ("Set A" measures) received time-endorsed status in 2010, thus supporting the content validity of these measures.

PINNACLE Registry: ACCF and AHA registries always attempt to include ACCF/AHA Task Force on Performance Measures in their various modules. The measures have content/context validity based on the approach articulated under the AACVPR/ACCF/AHA performance measure set.

Predictive Validity

Published data have shown that as the cardiac rehabilitation referral measure is met (i.e., the patient is referred to cardiac rehabilitation), the proximate desired outcome (cardiac rehabilitation participation) increases, as does the longer term desired outcome (reduction in morbidity and mortality rates). For more details, see the supplemental materials.

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

There were 17 individuals who completed the survey. Further information on the survey respondents is available if needed. Results of the survey were as follows:

-Average score: 4.12

-88.24% of respondents either agree or strongly agree that the outpatient measure can accurately distinguish good and poor quality.

AACVPR/ACCF/AHA performance measures set:

In May 2007 the final peer reviewed publication of the performance measures document was approved by the American Association of Cardiovascular and Pulmonary Rehabilitation Board of Directors, the American College of Cardiology Foundation Board of Trustees and by the American Heart Association Science Advisory and Coordinating Committee. Additionally, the publication was endorsed by the American College of Chest Physicians, American College of Sports Medicine, American Physical Therapy Association, Canadian Association of Cardiac Rehabilitation, European Association for Cardiovascular Prevention and Rehabilitation, Inter-American Heart Foundation, National Association of Clinical Nurse Specialists, Preventive Cardiovascular Nurses Association, and the Society of Thoracic Surgeons. The final document was published *the Journal of the American College of Cardiology* (the official journal of the American College of Cardiology), *the Journal of Cardiopulmonary Rehabilitation and Prevention* (the official journal of the American Association of Cardiovascular and Pulmonary Rehabilitation) and *Circulation* (the official journal of the American Heart Association) in September 2007. The document can be found at http://content.onlinejacc.org/cgi/reprint/j.jacc.2007.04.033v1.pdf.

AACVPR/ACCF/AHA CR3Project:

The cardiac rehabilitation referral measures (NQF measures 0642 and 0643) were revised in 2010 to clarify numerator and denominator exclusion criteria and to facilitate the implementation of these two measures by including administrative codes to identify denominator-eligible populations.

PINNACLE Registry:

A review of the measure based on the attributes, of reliability, ease of implementation, appropriate numerator, denominator, and exception specifications was performed. Given that it fulfilled these attributes, the measure was included in the registry. Data from this registry can be seen throughout the submission form and supplemental materials.

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

As noted above, our interpretation is that face and content validity has been established for this measure.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section* 2b4

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

AACVPR/ACCF/AHA CR3Project: Reliability of abstracting measure exclusions was tested in the CR3 project (see reliability testing section above). Exclusions or exceptions include patient, medical, and system-based conditions that would preclude the reasonable participation of a patient in a cardiac rehabilitation program (death, residing in an extended care nursing facility, lack of a cardiac rehabilitation program close to where the patient lives, etc.).

PINNACLE Registry: Exclusion rates and reasons were assessed from the PINNACLE Registry.

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

The reliability of abstracting exclusions was high to very high, as shown in the reliability section of this document.

In the PINNACLE Registry, 95% (n=944) of the providers did not report exclusions/exceptions. Among those providers who did report exclusions/exceptions, the mean rate was 29%. Among the patients with exclusions/exceptions, 7.4% were for medical reasons, 63.6 for patient reasons, and 29% for system reasons.

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

Documentation and assessment of exclusions are very important for this measure, to help reduce the possibility of bias in reporting (i.e., excluding patients who are actually eligible for cardiac rehabilitation referral, in order to improve performance scores). Based on the results of our CR3 project, the time and effort to assess exclusions does not appear to add significant burden (see supplemental materials section for more details on time required to complete abstraction of data for this measure).

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

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with _risk factors**

☐ **Stratification by _risk categories**

☐ **Other,**

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

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

Referral to cardiac rehabilitation is appropriate and evidence-based for all patients who have had a qualifying event/diagnosis/procedure. Referral or non-referral is not based on a patient's level of risk, but rather cardiac rehabilitation is appropriate and evidence-based for all eligible patients no matter what their risk level

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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

If stratified, skip to 2b3.9

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

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

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

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

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

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the*

measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

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

Performance rates were assessed by chart abstraction in the CR3 project. Reliability of that assessment was also performed, as noted in the reliability section above.

In the PINNACLE Registry, performance rates were assessed by decile, to allow for assessment of differences between "low" and "high" performing centers.

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

See reliability section for results of CR3 Project.

The PINNACLE Registry results are shown below:

# of providers	# of patients	Minimum	Lower Quartile	Mean	Upper Quartile	Maximum	Quartile Range	Std Dev
1022	298206	0.00%	0.84%	9.18%	13.0%	100%	12.1%	12.3%

	Mean
Decile 2	0.0%
Decile 3	0.9%
Decile 4	2.4%
Decile 5	4.1%
Decile 6	6.2%
Decile 7	8.9%
Decile 8	13.0%
Decile 9	19.0%
Decile 10	36.9%

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

There is wide variation in performance for this measure, documented in the datasets we used. Use of this measure allows for identification of that variation in delivery of cardiac rehabilitation referral. This is important because it provides data from which centers can identify improve upon gaps in care that are identified.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

In the CR3 Project, lack of documentation of a cardiac rehabilitation referral was assumed to represent "no referral made". In the PINNACLE database, missing values are interpreted as "no" responses. While it is challenging to ascertain a response that is truly "missing" versus one that is truly "No", we assume that data were missing if all records from a given practice are missing.

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

Given our assumptions, noted above, we did not conduct an empirical analysis of the frequency or distribution of missing data in the CR3 project.

In the PINNACLE dataset, 1.2% (13/1022) of centers were identified as having missing data and were excluded from analysis from the PINNACLE Registry.

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

Our assumption, based on the data listed above, is that the missing data rate is extremely low for our primary 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.

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

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

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

Some data elements are in defined fields in electronic sources

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

ACC is in the process of developing a common data dictionary mapped to coded terminology standards with the intent of improving interoperability with EHRs and potentially creation of emeasures

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

Attachment:

3c. Data Collection Strategy

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

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

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

ACC PINNACLE Registry:

PINNACLE project managers with clinical backgrounds provide significant support to local practices collecting data for PINNACLE.Unfortunately even with this support, experience has shown that clinicians still do not document this measure even if a patient has been referred to CR. It is hoped that CMS through its various initiatives will incentivize clinicians to both ensure eligible patients are being referred to a rehabilitation facility AND to be sure they are properly document the measure correctly.

ACC Cardiology Practice Improvement Pathway:

Although this program is now sunsetted, we learned it is hard to capture the difference between those already participating in CR vs. those referred to CR because of the long measurement window.

AACVPR/ACC/AHA CR3 Project:

The CR3 Project found that data abstraction of the CR/SP performance measure for referral from an inpatient setting is highly reliable, valid and feasible. However, we learned something about the definition of referral that will be a focus of future study and consideration.

CR/SP Referral is defined as including these 3 components:

1.) Documentation that patient was referred

2.) Communication (electronic/written) that referral information was given to patient

3.) Communication (electronic/written) that the receiving CR site was given patient's referral information.

Current practices and existing registries have typically only required the first component (i.e., any documentation that the patient was referred) in order to meet the performance measure (option 1). Because of this fact, we performed our reliability testing and predictive validity testing using this definition of referral. However, we recognize that the use of a stricter definition of referral that includes all 3 components listed above may increase the predictive validity of the measure (i.e., may increase the percentage of referred patients who enroll in CR/SP). Going forward, with the advent of better data collection systems we expect to be able to test the hypothesis that a stricter definition of CR/SP referral will increase enrollment in CR/SP.

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

None

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting Physician Compare https://www.medicare.gov/physiciancompare/ Payment Program MACRA/QPP/MIPs https://qpp.cms.gov/ MACRA/QPP/MIPs https://qpp.cms.gov/ Quality Improvement (external benchmarking to organizations) Quality Improvement with Benchmarking http://www.ncdr.com/webncdr/pinnacle/

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

PINNACLE Registry (URL: http://www.ncdr.com/webncdr/pinnacle/.
The PINNACLE Registry, part of the NCDR, is the largest ambulatory registry of its kind with over 26 million patient encounters from 8.9 million unique patients. It collects data from over 4,800 cardiologists, nurse practitioners (NPs), and physician assistants (PAs), largely using a system capable of directly extracting relevant information from electronic

health records, as referenced above. The primary purpose of the PINNACLE Registry is facilitating improvement in outpatient cardiovascular care quality and, by extension, improving patient outcomes. Utilizing established guidelines and performance measures, the PINNACLE Registry was designed to drive care improvement by reducing inappropriate variations in care, eliminating gaps in care, and improving care coordination for patients with cardiovascular diseases. The PINNACLE Registry assists practices in understanding and improving care through on demand performance reports for data-submitting practices and physicians. These reports, covering all valid patient encounters, detail adherence to over 30 cardiovascular clinical measures at the physician, location, and practice levels across coronary artery disease, hypertension, heart failure, and atrial fibrillation.
URL: https://qpp.cms.gov/ (Centers for Medicare and Medicaid Services):
The Merit-based Incentive Payment System (MIPS) is part of the quality payment program (QPP) which provides eligible Medicare professionals with a performance-based payment adjustment. In 2017, in order to be part of the QPP program, an eligible professionals would need to have billed Medicare more than $30,000 in Part B allowed charges a year and provide care for more than 100 Medicare patients a year. The MIPS payment adjustment for eligible professionals is determined on the data submitted as well as the duration of the data being submitted. It is our understanding that CMS is also planning to move towards publicly reporting physician data via Physician Compare.

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

Physicians who opt in to report on this measure for QPP/MIPS, are by extension agreeing for it to be reported on Physician Compare.
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 AACVPR 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.

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

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

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

Data are provided monthly via the PINNACLE Physician Dashboard to all PINNACLE participants. The dashboard provides a list of patients that met performance and did not meet performance on the measures. Algorithms and additional measure logic is available for the physician to review and understand how patients are captured. PINNACLE participants also work with client account managers to ensure that the data are being captured accurately from the electronic health record.

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

Data are provided monthly via the PINNACLE Physician Dashboard. The dashboard provides a list of patients that met performance and did not meet performance. Algorithm and additional measure logic is available for the physician to review and understand how patients are captured. PINNACLE participants work with client account managers that work with the practice to ensure that the data are being captured accurately from the electronic health record.

PINNACLE Participants also have access to Quality Improvement Toolkits available at the QII website.

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

Describe how feedback was obtained.

No feedback was obtained.

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

No feedback was obtained.

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

No feedback was obtained.

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

N/A

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

There does not appear to be a demonstrated improvement with a performance rate change of cardiac rehab referral of 5.5% in 2015 to 5.4% 2016. However, a study conducted on trends in referral to cardiac rehab after myocardial infarction showed a statistically significant increase in referral rates from 2007 to 2012(1). Furthermore, a study from the Centers for Disease Control and Prevention that looked at the use of outpatient cardiac rehabilitation among heart attack survivors in various states, found that after a heart attack, patients using cardiac rehab were 53% less likely to experience cardiac-related mortality than were those who did not use cardiac rehab (2). Based on the literature, we believe that continued implementation of the measure will lead to greater awareness and accountability among providers and accelerate improvements in referral (and enrollment) rates.

Citations:

(1) Beatty AL, Li S, Thomas L, et al. Trends in referral to cardiac rehabilitation after myocardial infarction: data from the National Cardiovascular Data Registry 2007 to 2012. J Am Coll Cardiol. 2014;63:2582-3.

(2) Fang J, Ayala C, Luncheon C, et al. Use of Outpatient Cardiac Rehabilitation Among Heart Attack Survivors - 20 States and the District of Columbia, 2013 and Four States, 2015. MMWR Morb Mortal Wkly Rep. 2017;66:869-73.

4b2. Unintended Consequences

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

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

No unintended negative consequences have been identified via our testing projects nor have any been reported to us by users of the measure.

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

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.

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?
5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

5b. Competing Measures

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

OR

Multiple measures are justified.

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

Appendix

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

Attachment Attachment: 0643_NQF_Submissions_Outpatient_Supplemental_Materials_2017-636456661317839252.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American College of Cardiology

Co.2 Point of Contact: Sana, Gokak, comment@acc.org, 202-375-6596-

Co.3 Measure Developer if different from Measure Steward: American College of Cardiology

Co.4 Point of Contact: Esteban, Perla, eperla@acc.org, 202-375-6499-

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.

Randal J. Thomas, MD, MS, FAHA, FACP, Chair; Marjorie King, MD, FACC, FAACVPR, member; Karen Lui, RN, C, MS, FAACVPR, member; Ileana L. Piña, MD, FACC, member;John Spertus, MD, MPH, FACC, member; Neil Oldridge, PhD, FAACVPR

The expert workgroup reviewed the available guidelines and other evidence, proposed and specified measures, responded to comments during peer review and public comment, continues to advise on additional specification of the measure and updates.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2007

Ad.3 Month and Year of most recent revision: 09, 2010

Ad.4 What is your frequency for review/update of this measure? Approximately every 3 years or as needed if evidence changes or due to feedback from implementation

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

Ad.6 Copyright statement: Copyright 2010, American Association for Cardiovascular and Pulmonary Rehabilitation, American College of Cardiology Foundation and the American Heart Association

Ad.7 Disclaimers: These measures and specifications are provided "as is" without warranty of any kind. Neither the AACVPR, the ACCF, nor the AHA shall be responsible for any use of these performance measures.

Limited proprietary coding is contained in the measure specifications (online data supplement) for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. The AACVPR, the ACCF, and the AHA disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT™) or other coding contained in the specifications.

CPT™ contained in the online data supplement is ©2009 American Medical Association.

Ad.8 Additional Information/Comments: None

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: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 3309

Corresponding Measures:

De.2. Measure Title: Risk-Standardized Survival Rate (RSSR) for In-Hospital Cardiac Arrest

Co.1.1. Measure Steward: American Heart Association

De.3. Brief Description of Measure: This measure estimates a hospital -level risk standardized survival rate (RSSR) for patients aged 18 years and older who experience an in-hospital cardiac arrest.

1b.1. Developer Rationale: Survival rates after in-hospital cardiac arrest vary across hospitals and serve as not only and indicator of patient severity of illness, but also as an indicator of success for the resuscitation structures and processes a facility has in place. To date, there has not been a risk-standardized survival rate measure for this population by which facilities can compare themselves to others. This measure is intended to fill that gap.

Chan PS, Berg RA, Spertus JA, Schwamm LH, Bhatt DL, Fonarow GC, et. al. Risk standardizing survivial for in-hospital cardiac arrest to facilitate hospital comparisons. JACC. 2013. 62:601-609.

S.4. Numerator Statement: Patients who were alive at discharge

S.6. Denominator Statement: Patients aged 18 years and older with in-hospital cardiac arrest who received chest compression and/or defibrillation

S.8. Denominator Exclusions: None

De.1. Measure Type: Outcome

S.17. Data Source: Registry Data

S.20. Level of Analysis: Facility

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

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable.

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

- The developer outlines several care processes that can be undertaken by the provider to influence patient survival at discharge, such as: the utilization of increased training of staff in resuscitation procedures (including the use of mock codes), earlier recognition of patients in cardiac arrest and shorter staff response time, and improved quality of chest compressions.
- The developer noted that survival rates post-in-hospital cardiac arrest have shown to improve with facility participation in the Get With The Guidelines-Resuscitation registry (from 16% up to 24% from 2010 to 2013) which could be linked to improved resuscitation care (Girota, et. al., 2012).

Question for the Committee:

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

Guidance from the Evidence Algorithm

Health outcome measure (Box 1) -> relationship between the measured health outcome and at least one healthcare action is demonstrated (Box 2)-> Pass

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

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

Maintenance measures – increased emphasis on gap and variation

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

- Based on a sample of 312 hospitals, the developer provides the following information:

Mean Performance Rate (Adjusted Survival)	**24%**
Median Performance Rate	24%
Standard Deviation	5%
Range of the Performance Rate	27%
Min, Max Rate	11%, 38%

- It is unclear which hospitals were included in this dataset. Characteristics of the measured hospitals and patients were not provided.

Disparities

- Race-specific survival was not assessed at the patient-level. The developer divided hospitals between 2011 and 2015 with at least 10 inpatient hospital cardiac arrest (IHCA) patients into quartiles of patients of black race. The median hospital percentage of IHCA patients of black race was 11% (IQR: 4% to 27%). Hospitals with the smallest number of black patients (quartile 1) had a higher unadjusted (observed) and RSSR for IHCA as compared with hospitals that had the highest number of black patients (quartile 4). The developer indicates that this data suggests some degree of disparity in RSSRs by hospital racial composition and therefore did not include race/ethnicity as a model covariate.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

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

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

1a. Evidence:

- This is a new measure with empirical data to support as well as a logic model to show processes leading to the desired outcome.
- Evidence is strong. The developer outlines several care processes that can be undertaken by the provider to influence patient survival at discharge, such as: the utilization of increased training of staff in resuscitation procedures (including the use of mock codes), earlier recognition of patients in cardiac arrest and shorter staff response time, and improved quality of chest compressions. The developer noted that survival rates post-in-hospital cardiac arrest have shown to improve with facility participation in the Get With The Guidelines-Resuscitation registry (from 16% up to 24% from 2010 to 2013) which could be linked to improved resuscitation care.
- The developer has demonstrated that survival can be improved with training.
- There is wide variation in survival rates after in-hospital cardiac arrest. This relates to multiple factors such as patient mix or to the recussitation efforts themselves. Comparing Hospital Survival Rates for Cardiac Arrest Using a Risk-Standardized Model Risk-Standardizing Survival for In-Hospital Cardiac Arrest to Facilitate HospitalvComparisons. Chan PS, Berg RA, Spertus JA, Schwamm LH, Bhatt DL, Fonarow GC, Heidenreich PA, Nallamothu BK, Tang F, Merchant RM. JACC
- This is a new measure and the evidence cited by the developer is related to clinical improvement in outcomes based on participating with the "Get with Guidelines" program. Although this program includes what are considered to be best practices with a cardiac arrest, this measure is comparing individual best practices of the program realtive to mortality. This is tangential evidence. To be direct, it would be comparing hospitals in the GWTG program with those that were not as numerator and survival rates as the denominator.
- Get with the guidelines has demonstrated the support for this measure.
- Moderate
- This measure is fatally flawed. The fundamental problem is that survival of an in-hospital cardiac arrest is conditional on having had an in-hospital cardiac arrest, and survivable in-hospital cardiac arrests are probably a marker of suboptimal care. Of the “cardiac arrests” on which this measure is based, more than 80% had a non-shockable rhythm (asystole or pulseless electrical activity). One can make a strong argument that any patient with asystole or PEA who survives an in-hospital arrest should never have arrested in the first place. If the arrest wasn’t preventable, the patient would not have survived. As hospital care approaches perfection, the survival rate for arrests other than VT/VF might therefore be expected to approach zero, not 100%. No amount of risk adjustment or hierarchical modeling can fix this problem.

- Several references to an association between the variability of certain processes of care and the survival of these patients are provided. It appears that these references are all on the GWTG dataset. Additionally, an ecological association between joining the GWTG and survival is suggested

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

- Racial disparity was identified.
- Survival rate is too low and the range is 11% to 38% (at least based on the relatively small sample of hospitals they studied). Hospitals with the smallest number of black patients (quartile 1) had a higher unadjusted (observed) and RSSR for IHCA as compared with hospitals that had the highest number of black patients suggesting racial disparities. I would rank opportunity for improvement high.
- Performance gap is 27 percentage points (11% to 38%). There is some evidence for disparities. Agree with moderate.
- There is a wide variation in outcomes across the institutions measured. This has been demonstrated in the JACC article.
- There is a performance gap identified in terms of survival that is significant.
- GWTG has shown a performance gap and responsiveness to interventions.
- Moderate
- N/A. Variability has been demonstrated, but there is nothing to tell us how this measure relates to actual quality of care.
- Adequate performance data are provided and demonstrate considerable variability in the GWTG population.

As one of the Scientific Acceptability reviewers commented, there are downsides to both including and excluding race from the adjustment. Stratification was suggested

Criteria 2: Scientific Acceptability of Measure Properties

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: Jeff Geppert, Matt Austin, Paul Kurlansky, Karen Joynt Maddox, David Cella

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

Review #1, Review #2, Review #3, Review #4, Review #5

Additional Information regarding Scientific Acceptability Evaluation (*if needed*): Reviewers were not able to come to consensus so both co-chairs also reviewed the measure.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Review #1: Scientific Acceptability

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.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 3309

Measure Title: Risk-Standardized Survival Rate (RSSR) for In-Hospital Cardiac Arrest

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
   NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.

   TIPS: Consider the following: 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?

   ☒Yes (go to Question #2)

   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

   The developer provides the data dictionary for the GWTG-Resuscitation registry. The numerator specification is discharge status alive [discharge disposition=2]. That implies that any value "not 2" for discharge disposition is "not alive" or "dead." Are there any other values for discharge disposition in the registry? In the UB-04, for example, there are values for transferred to another facility.

2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

   TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

   ☒Yes (go to Question #4)

   ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

3. Was **empirical VALIDITY testing** of patient-level data conducted?

   ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

   ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)

4. Was reliability testing conducted with computed performance measure scores for each measured entity?

   TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

   ☒Yes (go to Question #5)

   ☐No (go to Question #8)

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

   TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

   ☒Yes (go to Question #6)

   ☐No (please explain below then go to Question #8)

   Note that the developer tests reliability on pooled data over a 4.5-year time-period (01/2011-05/2015). Calculating the measure on a one-year time-period would result in reduced reliability at the average number of events, as well as reduce the number of hospitals with at least the minimum number of events (10).

6. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

   TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?

   Do the results demonstrate sufficient reliability so that differences in performance can be identified?

☐High (go to Question #8)

☒Moderate (go to Question #8)

☐Low (please explain below then go to Question #7)

7. Was other reliability testing reported?

   ☐Yes (go to Question #8)

   ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

8. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

   TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

   ☐Yes (go to Question #9)

   ☒No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)

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

   TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

   Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

   ☐Yes (go to Question #10)

   ☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

10. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

    ☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

    ☐Insufficient (go to Question #11)

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

1. Were all potential threats to validity that are relevant to the measure empirically assessed?

   TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

   ☐Yes (go to Question #2)

   ☒No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

   Distortion (i.e. behavior not aligned with the quality objective) attributable to the use of an in-hospital survival measure rather than an out-of-hospital survival measure (e.g. 30-day survival). Examples of distortion might include early discharge or transfers to non-acute facilities.

2. Analysis of potential threats to validity: Any concerns with measure exclusions?

   TIPS: Consider the following: 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)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

   ☒Yes (please explain below then go to Question #3)

   ☐No (go to Question #3)

   ☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

   Note that typical exclusions for an in-hospital mortality / survival measure might include:

   1) Patients with advanced directives in place prior to episode which specifically restrict any hospital specific protocol interventions or who decline (or their proxy declines) treatment; and (the developer does address this)
   2) Patients who have been transferred from one acute care hospital to another (the developer does not address this)

3. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

   ☐Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

   a. Is a conceptual rationale for social risk factors included? ☒Yes ☐No

   b. Are social risk factors included in risk model? ☐Yes ☒No

   c. Any concerns regarding the risk-adjustment approach?

   *TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? 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, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

   ☒Yes (please explain below then go to Question #4)

☐No (go to Question #4)

The developer rationale for not including social risk factors is two-fold. First, the GWTG-Resuscitation registry does not collect such variables (other than race). Second, that including social risk factors in the risk-adjustment model would "provide an exception to worse care for patients of lower socioeconomic status." The NQF Guidance does not specifically address the use of race as a social risk factor, but my understanding is that race is not to be used as a proxy for socioeconomic status and should only be included in the risk-adjustment model if there is an explicit clinical rationale. Per the NQF Guidance, "it is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences." The developer might have considered stratification rather than risk-adjustment. Finally, the developer focuses on only one of four possible causal pathways between a social risk factor and an outcome:

- A greater burden of disease when patient presents at the hospital
- More likely to use lower quality hospitals
- Patients may receive differential care in hospital
- Contextual factors independent of hospital quality
    - E.g. Competing economic priorities; access to post-acute care

With respect to model performance, the developer reports statistics on model discrimination (using the c-statistic) and calibration (using an R-squared statistic rather than the more conventional Hosmer-Lemeshow statistic).

Finally, the risk-adjustment model includes a variable for interventions that occur during the episode (e.g. mechanical ventilation, IV Vasopressor, and dialysis). A clinical expert might be able to determine whether any of these interventions might be correlated with "resuscitation structures and processes" and might therefore be confounded with the quality construct.

4. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

   ☒Yes (please explain below then go to Question #5)

   ☐No (go to Question #5)

   The developer does not report any statistics on the proportion of hospitals that are statistically better or worse than a threshold, benchmark, or median.

5. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

   ☐Yes (please explain below then go to Question #6)

   ☐No (go to Question #6)

   ☒Not applicable (go to Question #6)

   The developer does not report any testing results from another data source (e.g. claims data)

6. Analysis of potential threats to validity: Any concerns regarding missing data?

   ☐Yes (please explain below then go to Question #7)

   ☒No (go to Question #7)

ASSESSMENT OF MEASURE TESTING

7. Was <u>empirical</u> validity testing conducted using the measure as specified and appropriate statistical test?

   Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

☐Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

☒No (please explain below then go to Question #8)

8. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

   TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

   ☒Yes (go to Question #9)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

9. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

   ☒Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

   ☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

10. Was validity testing conducted with computed performance measure scores for each measured entity?

    TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

    ☐Yes (go to Question #11)

    ☐No (please explain below and go to Question #13)

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

    TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

    ☐Yes (go to Question #12)

    ☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

12. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #14)

    ☐Moderate (go to Question #14)

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

13. Was other validity testing reported?

    ☐Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

14. Was validity testing conducted with patient-level data elements?

TIPS: Prior validity studies of the same data elements may be submitted

☐Yes (go to Question #15)

☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

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

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #16)

    ☐No (please explain below and rate Question #16 as INSUFFICIENT)

16. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

    ☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

    ☐Insufficient (go to Question #17)

17. OVERALL VALIDITY RATING

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

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

The 1.b.1 Developer Rationale suggests an empirical validity test that might be conducted based on an association with "resuscitation structures and processes." The developer might consider collected data on these structures and processes and conducting such an empirical validity test. The developer cites other potential threats to validity in the peer-reviewed journal article (Chan et. al.) including unmeasured confounders, facility variation in documentation, and the absence of evidence on reliability and validity for hospitals that do not participate in the GWTG-Resuscitation registry.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

Review #2: Scientific Acceptability

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.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 3309

Measure Title: Risk-Standardized Survival Rate for In-Hospital Cardiac Arrest

RELIABILITY

11. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?

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

    TIPS: Consider the following: 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?

    ☒Yes (go to Question #2)

    ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

12. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

    TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

    ☒Yes (go to Question #4)

    ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

13. Was **empirical VALIDITY testing** of patient-level data conducted?

    ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

    ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)

14. Was reliability testing conducted with computed performance measure scores for each measured entity?

TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

☒Yes (go to Question #5)

☐No (go to Question #8)

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

    TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

    ☒Yes (go to Question #6)

    ☐No (please explain below then go to Question #8)

16. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?

    Do the results demonstrate sufficient reliability so that differences in performance can be identified?

    ☐High (go to Question #8)

    ☒Moderate (go to Question #8)

    ☐Low (please explain below then go to Question #7)

17. Was other reliability testing reported?

    ☐Yes (go to Question #8)

    ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

18. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

    TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

    ☐Yes (go to Question #9)

    ☒No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)

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

    TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

    Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #10)

    ☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

20. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

    ☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

☐Insufficient (go to Question #11)

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

17. Were all potential threats to validity that are relevant to the measure empirically assessed?

TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

☒Yes (go to Question #2)

☐No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

18. Analysis of potential threats to validity: Any concerns with measure exclusions?

TIPS: Consider the following: 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)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

☐Yes (please explain below then go to Question #3)

☒No (go to Question #3)

☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

19. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

☐Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

a. Is a conceptual rationale for social risk factors included? ☒Yes ☐No

b. Are social risk factors included in risk model? ☐Yes ☒No

c. Any concerns regarding the risk-adjustment approach?

*TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? 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, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g.,*

adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?

☐Yes (please explain below then go to Question #4)

☒No (go to Question #4)

20. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

☐Yes (please explain below then go to Question #5)

☒No (go to Question #5)

21. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

☐Yes (please explain below then go to Question #6)

☐No (go to Question #6)

☒Not applicable (go to Question #6)

22. Analysis of potential threats to validity: Any concerns regarding missing data?

☒Yes (please explain below then go to Question #7)

☐No (go to Question #7)

The impact of missing data was not assessed and it is not clear why it was not. They referenced that 14 hospitals did not have complete data and were excluded from the testing sample.

ASSESSMENT OF MEASURE TESTING

23. Was empirical validity testing conducted using the measure as specified and appropriate statistical test?

Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

☐Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

☒No (please explain below then go to Question #8)

Chose to use face validity.

24. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

☒Yes (go to Question #9)

☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

25. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

☒Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as

INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

26. Was validity testing conducted with computed performance measure scores for each measured entity?

    TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

    ☐Yes (go to Question #11)

    ☐No (please explain below and go to Question #13)

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

    TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

    ☐Yes (go to Question #12)

    ☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

28. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #14)

    ☐Moderate (go to Question #14)

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

29. Was other validity testing reported?

    ☐Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

30. Was validity testing conducted with patient-level data elements?

    TIPS: Prior validity studies of the same data elements may be submitted

    ☐Yes (go to Question #15)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

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

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #16)

    ☐No (please explain below and rate Question #16 as INSUFFICIENT)

32. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

☐Insufficient (go to Question #17)

17. OVERALL VALIDITY RATING

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

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

Review #3: Scientific Acceptability

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.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 3309

Measure Title: Risk-Standardized Survival Rate (RSSR) for In-Hospital Cardiac Arrest

RELIABILITY

21. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
    NOTE: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.

    TIPS: Consider the following: 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?

    ☒Yes (go to Question #2)

    ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

22. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

    TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

    ☒Yes (go to Question #4)

    ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

23. Was **empirical VALIDITY testing** of patient-level data conducted?

    ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

    ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)

24. Was reliability testing conducted with computed performance measure scores for each measured entity?

    TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

    ☒Yes (go to Question #5)

    ☐No (go to Question #8)

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

    TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

    ☒Yes (go to Question #6)

    ☐No (please explain below then go to Question #8)

26. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?

    Do the results demonstrate sufficient reliability so that differences in performance can be identified?

    ☐High (go to Question #8)

    ☒Moderate (go to Question #8)

    ☐Low (please explain below then go to Question #7)

27. Was other reliability testing reported?

    ☐Yes (go to Question #8)

☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

28. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

    TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

    ☐Yes (go to Question #9)

    ☒No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)

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

    TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

    Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #10)

    ☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

30. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

    ☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

    ☐Insufficient (go to Question #11)

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

Their reliability testing was rather sophisticated, looking at fixed and random effects to parse out within hospital noise from signal. I don't understand why there was so much noise within hospital that the reliability coefficient is a mere 0.70. I assume the error is in measurement of risk adjusting covariates. I believe that reliability of patient-level data should be high, assuming chart data being abstracted are accurate, but these data came from the registry.

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

33. Were all potential threats to validity that are relevant to the measure empirically assessed?

TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

☐Yes (go to Question #2)

☒No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

It's not possible to empirically assess all threats to validity. 30-day post discharge would remove a major threat to the threat posed by early hospital discharge. Regarding generalizability outside of the GWGR Registry, I think now the AHA is proposing that this first PM be applied only to registry participants?

34. Analysis of potential threats to validity: Any concerns with measure exclusions?

TIPS: Consider the following: 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)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

☒Yes (please explain below then go to Question #3)

☐No (go to Question #3)

☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

Good point that patients with advance directive for no CPR should be excluded, and the threat posed by d/c to another acute facility arguing for 30-day f/u.

35. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

☐Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

a. Is a conceptual rationale for social risk factors included? ☒Yes ☐No

b. Are social risk factors included in risk model? ☐Yes ☒No

c. Any concerns regarding the risk-adjustment approach?

*TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? 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, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

☒Yes (please explain below then go to Question #4)

☐No (go to Question #4)

I was persuaded by the decision not to include race (given higher risk in blacks). It seems unintended consequences could result from either choice (exclusion or inclusion).

36. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

☐Yes (please explain below then go to Question #5)

☒No (go to Question #5)

37. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

☐Yes (please explain below then go to Question #6)

☐No (go to Question #6)

☒Not applicable (go to Question #6)

38. Analysis of potential threats to validity: Any concerns regarding missing data?

☐Yes (please explain below then go to Question #7)

☒No (go to Question #7)

There was very little missing data.

ASSESSMENT OF MEASURE TESTING

39. Was empirical validity testing conducted using the measure as specified and appropriate statistical test?

Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

☐Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

☒No (please explain below then go to Question #8)

40. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

☒Yes (go to Question #9)

☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

41. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

☒Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

42. Was validity testing conducted with computed performance measure scores for each measured entity?

TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

☐Yes (go to Question #11)

☐No (please explain below and go to Question #13)

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

TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

☐Yes (go to Question #12)

☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

44. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #14)

    ☐Moderate (go to Question #14)

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

45. Was other validity testing reported?

    ☐Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

46. Was validity testing conducted with patient-level data elements?

    TIPS: Prior validity studies of the same data elements may be submitted

    ☐Yes (go to Question #15)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

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

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #16)

    ☐No (please explain below and rate Question #16 as INSUFFICIENT)

48. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

    ☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

    ☐Insufficient (go to Question #17)

17. OVERALL VALIDITY RATING

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

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

My bottom line is that this measure should go forward to committee for deliberation. With a generally favorable methods committee approval, although the caveats should help the parent committee in those deliberations.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

Review #4: Scientific Acceptability

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.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 3309

Measure Title: Risk-Standardized Survival Rate (RSSR) for In-Hospital Cardiac Arrest

RELIABILITY

31. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?

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

    TIPS: Consider the following: 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?

    ☒Yes (go to Question #2)

    ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

32. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

    TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

    ☒Yes (go to Question #4)

    ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

33. Was **empirical VALIDITY testing** of patient-level data conducted?

    ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

    ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION)

34. Was reliability testing conducted with computed performance measure scores for each measured entity?

TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

☒Yes (go to Question #5)

☐No (go to Question #8)

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

    TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

    ☒Yes (go to Question #6)

    ☐No (please explain below then go to Question #8)

36. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?

    Do the results demonstrate sufficient reliability so that differences in performance can be identified?

    ☐High (go to Question #8)

    ☒Moderate (go to Question #8)

    ☐Low (please explain below then go to Question #7)

37. Was other reliability testing reported?

    ☐Yes (go to Question #8)

    ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

38. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

    TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

    ☐Yes (go to Question #9)

    ☒No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION)

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

    TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

    Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #10)

    ☐No (if no, please explain below and rate Question #10 as INSUFFICIENT)

40. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

    ☐Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

    ☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

☐Insufficient (go to Question #11)

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

49. Were all potential threats to validity that are relevant to the measure empirically assessed?

TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

☒Yes (go to Question #2)

☐No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

50. Analysis of potential threats to validity: Any concerns with measure exclusions?

TIPS: Consider the following: 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)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

☐Yes (please explain below then go to Question #3)

☒No (go to Question #3)

☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

51. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

☐Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

a. Is a conceptual rationale for social risk factors included? ☒Yes ☐No

b. Are social risk factors included in risk model? ☐Yes ☒No

c. Any concerns regarding the risk-adjustment approach?

*TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? 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, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g.,*

adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?

☒Yes (please explain below then go to Question #4)

☐No (go to Question #4)

One potential issue with hierarchical models is that if there is a volume effect (i.e. high-volume hospitals do better) this is lost in the modeling technique. This is something that (in my opinion) should always be tested when hierarchical models are submitted – need to see raw and adjusted performance by quintiles or deciles of volume. This may be an interesting topic for discussion at one of our calls since it applies essentially to all risk-standardized measures.

52. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

☒Yes (please explain below then go to Question #5)

☐No (go to Question #5)

I'm not sure if this is an appropriate concern, but the participants in the GWTG registry are a highly selected group of hospitals electing to participate in quality improvement efforts. If this were to be expanded beyond this registry, I'm not sure any of the testing here would apply.

53. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

☐Yes (please explain below then go to Question #6)

☒No (go to Question #6)

☐Not applicable (go to Question #6)

54. Analysis of potential threats to validity: Any concerns regarding missing data?

☐Yes (please explain below then go to Question #7)

☒No (go to Question #7)

ASSESSMENT OF MEASURE TESTING

55. Was <u>empirical</u> validity testing conducted using the measure as specified and appropriate statistical test?

Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

☐Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

☒No (please explain below then go to Question #8)

Not provided

56. Was <u>face validity</u> systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

☒Yes (go to Question #9)

☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

57. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

☒Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

The experts seemed to be in general agreement that the concept has face validity. More information on the degree of consensus and the areas of disagreement would be of use in evaluating this concept more fully.

58. Was validity testing conducted with computed performance measure scores for each measured entity?

TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

☐Yes (go to Question #11)

☐No (please explain below and go to Question #13)

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

TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

☐Yes (go to Question #12)

☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

60. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

☐High (go to Question #14)

☐Moderate (go to Question #14)

☐Low (please explain below then go to Question #13)

☐Insufficient

61. Was other validity testing reported?

☐Yes (go to Question #14)

☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

62. Was validity testing conducted with patient-level data elements?

TIPS: Prior validity studies of the same data elements may be submitted

☐Yes (go to Question #15)

☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on score-level rating from Question #12)

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

TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

☐Yes (go to Question #16)

☐No (please explain below and rate Question #16 as INSUFFICIENT)

64. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

☐Insufficient (go to Question #17)

17. OVERALL VALIDITY RATING

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

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

Only face validity testing has been performed, but as a new measure that meets moderate validity criteria.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

Review #5: Scientific Acceptability

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.**

Instructions:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions.
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- We have provided TIPS to help you answer the questions.
- We've designed this form to try to minimize the amount of writing that you have to do. That said, ***it is critical that you explain your thinking/rationale if you check boxes where we ask for an explanation*** (because this is a Word document, you can just add your explanation below the checkbox)**.** Feel free to add additional explanation, even if an explanation is not requested (but please type this underneath the appropriate checkbox).
- This form is based on Algorithms 2 and 3 in the Measure Evaluation Criteria and Guidance document (see pages 18-24). These algorithms provide guidance to help you rate the Reliability and Validity subcriteria. ***We ask that you refer to this document when you are evaluating your measures***.
- Please contact Methods Panel staff if you have questions (methodspanel@qualityforum.org).

Measure Number: 3309

Measure Title: Risk-Standardized Survival Rate (RSSR) for In-Hospital Cardiac Arrest

RELIABILITY

41. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?

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

    TIPS: Consider the following: 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?

    ☒Yes (go to Question #2)

    ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

42. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

    TIPS: Check the 2nd "NO" box below if: only descriptive statistics provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level of analysis, patients)

    ☒Yes (go to Question #4) Yes

    ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified OR there is no reliability testing (please explain below then go to Question #3)

43. Was **empirical VALIDITY testing** of patient-level data conducted?

    ☐Yes (use your rating from data element validity testing – Question #16- under Validity Section)

    ☐No (please explain below and rate Question #11: OVERALL RELIABILITY as INSUFFICIENT and proceed to the VALIDITY SECTION) No; Face validity of the metric as a metric of quality for hospital level comparison was performed

44. Was reliability testing conducted with computed performance measure scores for each measured entity?

    TIPS: Answer no if: only one overall score for all patients in sample used for testing patient-level data

    ☒Yes (go to Question #5)

    ☐No (go to Question #8) No

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

    TIPS: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

    ☒Yes (go to Question #6)

    ☐No (please explain below then go to Question #8)

46. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

    TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation?

    Do the results demonstrate sufficient reliability so that differences in performance can be identified?

    ☐High (go to Question #8)

    ☒Moderate (go to Question #8)

    ☐Low (please explain below then go to Question #7)

47. Was other reliability testing reported?

    ☐Yes (go to Question #8) Yes

    ☐No (rate Question #11: OVERALL RELIABILITY as LOW and proceed to the VALIDITY SECTION)

48. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

    TIPS: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" see Validity Section Question #15)

    ☐Yes (go to Question #9)

    ☒No (if there is score-level testing, rate Question #11: OVERALL RELIABILITY based on score-level rating from Question #6; otherwise, rate Question #11: OVERALL RELIABILITY as INSUFFICIENT. Then proceed to the VALIDITY SECTION) Score level testing; reliability moderate

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

    TIPS: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

    Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #10)

    ☐No (if no, please explain below and rate Question #10 as INSUFFICIENT) Reliability testing was performed for the score. Data elements are not measures but statement of fact—e.g. age, sex, history of prior heart failure, pulseless electrical activity, etc. Outcome is patient alive at hospital discharge. All of these elements are clearly defined and require only correct input of the information from the hospital record but do not involve observational interpretation (as would, for example, reading of an echocardiogram or radiology study). Audit of the database upon which this metric is based has been published and documents the reliability of the data itself. Moreover, audit has revealed the extremely low percentage of missing data—mechanism for management has been explained to NQF and is methodologically appropriate.

50. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

TIPS: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?

☐Moderate (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as MODERATE)

☐Low (if score-level testing was NOT conducted, rate Question #11: OVERALL RELIABILITY as LOW)

☐Insufficient (go to Question #11) Based on comments from #9 above, would have to rate data element reliability as insufficient—it may in fact be sufficient but evidence to confirm this has not been presented.

11. OVERALL RELIABILITY RATING

OVERALL RATING OF RELIABILITY taking into account precision of specifications and all testing results:

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

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

☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

65. Were all potential threats to validity that are relevant to the measure empirically assessed?

TIPS: Threats to validity include: exclusions; need for risk adjustment; Able to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.

☐Yes (go to Question #2)

☒No (please explain below and go to Question #2) [NOTE that even if ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***, we still want you to look at the testing results]

There are two potential threats to validity—one to internal validity, the other to external validity: 1) exclusion of race as a data element which is addressed in #3 below; 2) concern with external validity—This measure was developed from data from 272 of the >300 hospitals who participate in the American Heart Association Get With the Guidelines: Resuscitation registry. Although this is an important registry, we have no idea how well the data from these hospitals reflects performance in general. Indeed, there is theoretical reason to believe that those sites who participate may be those that are more motivated and interested in the care of these patients and actually tend to perform better than sites that don't. Therefore as a metric to compare performance among those hospitals that participate in the registry the metric may have considerable validity—among those that don't, we have no idea. Moreover, from a functional point of view, a hospital would need to be participating in the registry in order to be rated. Therefore, we cannot say that this measure has general validity as a measure of the quality of hospitals' performance and, although not an issue of validity but rather applicability, there is no mechanism to use this measure for the vast majority of hospitals that do not participate. Perhaps, within the limited scope of presentation as a measure specifically for those participating in the GWTG registry, this concern regarding "external" validity need not enter the discussion. However, should NQF endorse the measure, the language of that endorsement should clearly note the limitation.

66. Analysis of potential threats to validity: Any concerns with measure exclusions?

TIPS: Consider the following: 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)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?

☐Yes (please explain below then go to Question #3)

☐No (go to Question #3)

☒Not applicable (i.e., there are no exclusions specified for the measure; go to Question #3)

67. Analysis of potential threats to validity: Risk-adjustment (applies to all outcome, cost, and resource use measures; may also apply to other types of measure)

☐Not applicable (e.g., structure or process measure that is not risk-adjusted; go to Question #4)

a. Is a conceptual rationale for social risk factors included? ☒Yes ☐No

b. Are social risk factors included in risk model? ☐Yes ☒No

c. Any concerns regarding the risk-adjustment approach?

*TIPS: Consider the following: If a justification for **not risk adjusting** is provided, is there any evidence that contradicts the developer's rationale and analysis? If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? **If risk adjusted**: Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? 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, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

☒Yes (please explain below then go to Question #4)

Risk-adjustment approach was generally excellent with one notable exception.

Metric authors have made an intentional decision to exclude race as a factor in their model. The rationale supplied is strange—that there were a higher proportion of one race in hospitals with poorer outcome. Same rationale could a priori be used to exclude any of the risk factors found to be significant in the model. In fact, decision appears to be more political than scientific in origin. Based on the published description of the risk model (J Am Coll Cardiol, 2013;62:601-609), decision was based on two factors: 1) AHA Scientific Statement of Standards for Statistical Models Used for Public Reporting of Health Outcomes, which states as a matter of opinion without validation that "the use of variables that convey nonclinical information(e.g. race/ethnicity, socioeconomic status) should generally be avoided"—ignoring the potential fact that racial differences may have clinical relevance. 2)The earlier work of Dr Chan et al on "Racial difference in survival after in-hospital cardiac arrest" (JAMA 2009;302:1195-1201) in which they concluded that "much of the racial difference in survival was associated with the hospital center in which black patients received care." However, their data actually showed that even after adjusting for age, sex, clinical characteristics and hospital, nearly 40% of the racial discrepancy remained unexplained. In their discussion, among the factors that might account for the residual variance they did entertain the possibility that there may be racial differences in the physiological response to cardiac arrest. In short, rather than taking the methodologically more sound approach of entering race into the model to see what, if any predictive value it might contribute, measure developers made an a priori decision to exclude. Although the rationale for this is concern that hospitals who care for black patients might be "let off the hook" for otherwise poor performance, similar exclusion of racial and socioeconomic factors in the readmission metric adopted by CMS based on the work of the same study group has resulted in the unintended consequence of penalizing safety net hospitals in greatest need of resource support.

In summary, regarding exclusion of race, 1) rationale given in metric papers was not rational; 2) decision was not scientifically based nor methodologically sound; and, 3) decision may actually have opposite of intended effect.

☐No (go to Question #4) Aside from concerns regarding measure exclusion, risk-adjustment approach was generally excellent.

68. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

☐Yes (please explain below then go to Question #5)

☒No (go to Question #5)Model appears to perform well

69. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

☐Yes (please explain below then go to Question #6)

☐No (go to Question #6)

☒Not applicable (go to Question #6)Not applicable

70. Analysis of potential threats to validity: Any concerns regarding missing data?

☐Yes (please explain below then go to Question #7)Yes

☒No (go to Question #7)

As noted above, there was apparently no missing data in the 312/326 sites used to test the model. Authors have subsequent to the application provided information regarding an appropriate strategy for management of missing data.

ASSESSMENT OF MEASURE TESTING

71. Was empirical validity testing conducted using the measure as specified and appropriate statistical test?

Answer no if: face validity; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).

☐Yes (go to Question #10) [NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary. Go to Question #8 **only if** there is insufficient information provided to evaluate data element and score-level testing.]

☒No (please explain below then go to Question #8)

72. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

TIPS: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.

☒Yes (go to Question #9)

☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT)

73. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

☒Yes (if a NEW measure, rate Question #17: OVERALL VALIDITY as MODERATE)

☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, rate Question #17: OVERALL VALIDITY as

INSUFFICIENT; otherwise, rate Question #17: OVERALL VALIDITY as MODERATE)

☐No (please explain below and rate Question #17: OVERALL VALIDITY AS LOW)

74. Was validity testing conducted with computed performance measure scores for each measured entity?

    TIPS: Answer no if: one overall score for all patients in sample used for testing patient-level data.

    ☐Yes (go to Question #11)

    ☐No (please explain below and go to Question #13)

75. Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?

    TIPS: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

    ☐Yes (go to Question #12)

    ☐No (please explain below, rate Question #12 as INSUFFICIENT and then go to Question #14)

76. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #14)

    ☐Moderate (go to Question #14)

    In absence of information regarding hospitals other than the 312 participants in the AHA Get With the Guidelines Resuscitation registry, the information is insufficient to conclude that the measure provides a valid measure of comparative hospital performance in the treatment of patients with in-hospital cardiac arrest. It does, however, provide a valid indicator of quality among hospitals who participate in the registry.

    ☐Low (please explain below then go to Question #13)

    ☐Insufficient

77. Was other validity testing reported?

    ☐Yes (go to Question #14)

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as LOW)

78. Was validity testing conducted with patient-level data elements?

    TIPS: Prior validity studies of the same data elements may be submitted

    ☐Yes (go to Question #15)Yes—model which was developed from data from patients in the registry from 2007-2010 using a derivation and test grouping approach; model was subsequently tested using 2011-2015 data and found to have excellent calibration

    ☐No (please explain below and rate Question #17: OVERALL VALIDITY as INSUFFICIENT if no

    score-level testing was conducted, otherwise, rate Question #17: OVERALL VALIDITY based on

    score-level rating from Question #12)

    Registry values were assumed to be correct and were the only source of data for developing the models developed. There was no testing of the patient level data from the medical records.

79. Was the method described and appropriate for assessing the accuracy of ALL critical data elements? *NOTE that data element validation from the literature is acceptable.*

    TIPS: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.

Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

☐Yes (go to Question #16) Yes

☐No (please explain below and rate Question #16 as INSUFFICIENT)

80. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

☐Moderate (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as MODERATE)

Concerns regarding representativeness of hospitals from which the data was drawn are outlined above in #1.

☐Low (please explain below) (if score-level testing was NOT conducted, rate Question #17: OVERALL VALIDITY as LOW)

☐Insufficient (go to Question #17

17. OVERALL VALIDITY RATING

OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.

☐High (NOTE: Can be HIGH only if score-level testing has been conducted)

☐Moderate (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☒Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required] [NOTE: If rating is INSUFFICIENT for all empirical testing, then go back to Question #8 and evaluate any face validity that was conducted, then reconsider this overall rating.]

Based on the concerns outlined in#3 above, regarding the a priori exclusion of race despite the evidence that it may provide additional meaningful information, construction of the metric does not appear to be optimal.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?

TIPS: Consider the following: Do the component measures fit the quality construct? Are the objectives of parsimony and simplicity achieved while supporting the quality construct?

☐High

☐Moderate

☐Low (please explain below)

☐Insufficient (please explain below)

Committee Pre-evaluation Comments: Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Specifications:

- They provided reasonable specifications that are captured in GWTG-Res. But why did they limit metrics to chest compressions and/or defibrillation? Why not include other metrics? Also what about pts who are transferred to other facilities?

- Reliability is moderate. One statistical reviewer questioned the absence of a "transferred" category for disposition
- The data elements can generally all be recognized and monitored. These include age group; initial cardiac arrest rhythm; hospital location of arrest; the presence of hypotension (abnormally low blood pressure), sepsis, metastatic or hematologic malignancy, or hepatic insufficiency prior to cardiac arrest; or requirement of mechanical ventilation of continuous vasopressor infusion at the time of cardiac arrest.
- The data elements are vague and qualitative (like running mock codes) and this will make the reliability of the measure poor.
- Moderate reliability
- Moderate
- Data elements are all present in GWTG. Whether would be standardized and generalizable to other sites is a question.

2a2. Reliability testing:

- Signal to noise ratio was 0.7, and didn't change when comparing average number of events to the minimum number of events.
- Moderate. They stated that their signal-to-noise ratio testing was conducted by fitting a hierarchical, logistic regression model to derive the two shape parameters – alpha and beta; the model was built on a specified beta-binomial distribution. This modeling was done on patient-level data, adjusting for age, gender, location of arrhythmia (i.e. ICU, ED), type of heart rhythm, and present on arrival (POA) conditions. This methodological approach enabled them to account for patient-level and hospital-level (random effects) mixed effects.
- No concerns. "Moderate" appears to be appropriate.
- None
- Yes, the actual numerator is vague unless it was to be hospitals participating in the GWTG program.
- Low testing reliability
- I am either misreading - or very confused.
- First, on p 54, the developers make several statements that do not seem to be consistent:

  A total of 326 hospitals reported on this measure. Of these, 312 hospitals had all the required data elements and met the minimum number of quality reporting events (1) for inclusion the analyses. For this measure, 96 percent of hospitals were included in the analyses, and the average number of quality reporting events was 190, for a total of 61,934 cardiac arrest events, and 14,782 cardiac arrest survivals to discharge. Of the 312 hospitals, the range of cardiac arrest quality reporting events was 1 to 122, and the range for survivals was 0 to 344.

  If the average was 190, how could the range be 1 to 122; and how could the range for survivals be 0 to 344?

  On page 58 they present a frequency distribution of ratings, from 1 to 5, but claim the mean is 6.8.

  I am also confused in that all of the scientific acceptability reviews responded 'NO" to "was reliability testing conducted with patient-level data elements..." It is no clear to me how the measure can be calculated without these elements. The hieraarchical model on p. 57 is not described well in terms of where the individual risk factors fit into the equation. One reviewer commented on the fact that 5 years of data were used to test reliability. Although plots are provided demonstrating good correlation between observed and expected, there remains the question of sensitivity with one year of data

2b2. Validity testing & 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data):

- Moderately high face validity.
- Moderate validity testing as they surveyed experts in the field and 71% either agreed or strongly agreed the measure is valid.
- Validity is "moderate".

- The missing data could have an impact upon the results but this is being calculated at such a high level that individual data elements are not likely to be included.
- I think validity testing will be difficult due to the lack of standardization between facilities.
- Moderate validity
- Low validity
- Two reviewers commented on the possibility of distortion such as premature discharge when evaluating in-hospital events. Only GWTG data have been used - both for evidence and for testing

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment):

- Does not include race in the risk adjustment mode.
- A number of demographic (age category, sex) and comorbidity variables (includes pre-existing conditions and interventions in place at the time of cardiac arrest) were considered for model inclusion. They consider almost all variables (except race) as potential predictors in the model. The table below lists all covariates (also termed predictors or risk factors) included in the full model, along with their corresponding estimates (also termed coefficients), their odds ratios (amount of risk relative to the reference population), and their 95% confidence intervals.
- Since this is an in-hospital measure, social concerns should not influence the outcomes.
- The primary areas of risk adjustment that may need attention have to do with its own data elements: age group; initial cardiac arrest rhythm; hospital location of arrest; the presence of hypotension (abnormally low blood pressure), sepsis, metastatic or hematologic malignancy, or hepatic insufficiency prior to cardiac arrest; or requirement of mechanical ventilation of continuous vasopressor infusion at the time of cardiac arrest.
- Risk adjustment is poor. Ethnicity/Race seems to be considered as an afterthought analysis rather than a discrete data field.
- Yes
- Risk adjustment is irrelevant here since the measure does not have face validity, but if the measure were not invalid I would agree with decision not to "adjust away" racial/socioeconomic differences.
- The risk-adjustment process is appropriate, although one reviewer suggested a potential volume relationship that was not explored. As mentioned above, the risk adjustment equation was not explicit in terms of the risk adjustment factors.

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent
3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than person obtaining original information

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?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments: Criteria 3: Feasibility

3. Feasibility:

- Appears feasible as data are available in GWTG-Res.
- All measure elements are readily available in electronic sources via administrative claims data, and coded by someone other than person obtaining original information. Reliability is "moderate"
- Feasibility of this measure is poor due to the inconsistency with which data can be recorded from facility to facility.
- Demonstrated by GWTG.
- Moderate feasibility
- All data elements are routinely collected in GWTG. However, it is not clear how generalizable and standardized the data elements are across sites that are not members of GWTG.

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details:

- American Heart Association Get With The Guidelines-Resuscitation Professional Certification or Recognition Program
    - Hospitals that participate actively and consistently in Get With The Guidelines®--Resuscitation are eligible for public recognition.
    - Awards recognize hospitals that demonstrate at least 85 percent compliance in each of the four Get With The Guidelines-Resuscitation Recognition Measures. The different levels reflect the amount of time for which the hospital demonstrates performance.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The measure and its specifications and results were vetted with the American Heart Association Research Committee chairs and feedback was provided. This feedback was incorporated into the final measure.
- Participants in the Get With The Guidelines-Resuscitation program have access to their data through the registry (also called the Patient Management Tool), where they are able to query and review results. Additionally, they

receive a separate feedback report, available as a pdf download, of their risk-standardized in-hospital cardiac arrest results.

Additional Feedback: None

***Questions for the Committee**:*

- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *How has the measure been vetted in real-world settings by those being measured or others?*

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results: The developer states that survival rates after in-hospital cardiac arrest had started to improve prior to the introduction of the feedback reports regarding results on the risk-standardized in-hospital cardiac arrest survival.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation: The developer does not list any unexpected findings.

Potential harms The developer does not list any unexpected findings.

***Questions for the Committee**:*

- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for Usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments: Criteria 4: Usability and Use

4a. Use:

- Not publicly reported? Current use in an accountability program: being used in GWTG-Res.
- The measure is currently used in the Get With the Guidelines Program oHospitals that participate actively and consistently in Get With The Guidelines®--Resuscitation are eligible for public recognition.
- This data would ultimately be very helpful in understanding the impacts of the various elements within the program.
- Currently not publicly reported. Only accountability is participation/recognition in GWTG program.
- Has been adopted by a large number of institutions.
- Moderate
- It is not clear that the specific measure is being used, although it is incorporated into GWTG.

4b. Usability:

- Use for quality improvement.
- Moderate usability.
- The measure appears to be usable.
- This is currently in use so it does have elements of usability

- Usability is only as it relates to participation in the GWTG program since chest compressions and defibrillation are going to be universal in the setting of cardiac arrest.
- Can be used to improve structure and processes that improved outcomes.
- Moderate
- For sites that collect the data, the measure is easily calculated. Whether results from GWTG would apply is another question.

Criterion 5: Related and Competing Measures

Related or competing measures
None

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 10, 2018

No comments have been submitted as of this date.

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

AHA-RSSR_Evidence_Attachment_v3.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission? Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

1a Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:**

Measure Title: Risk-Standardized Survival Rate (RSSR) for In-Hospital Cardiac Arrest

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 11/8/2017

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.

- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Outcome: Patient Survival at Discharge

☐Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

☐ Process:

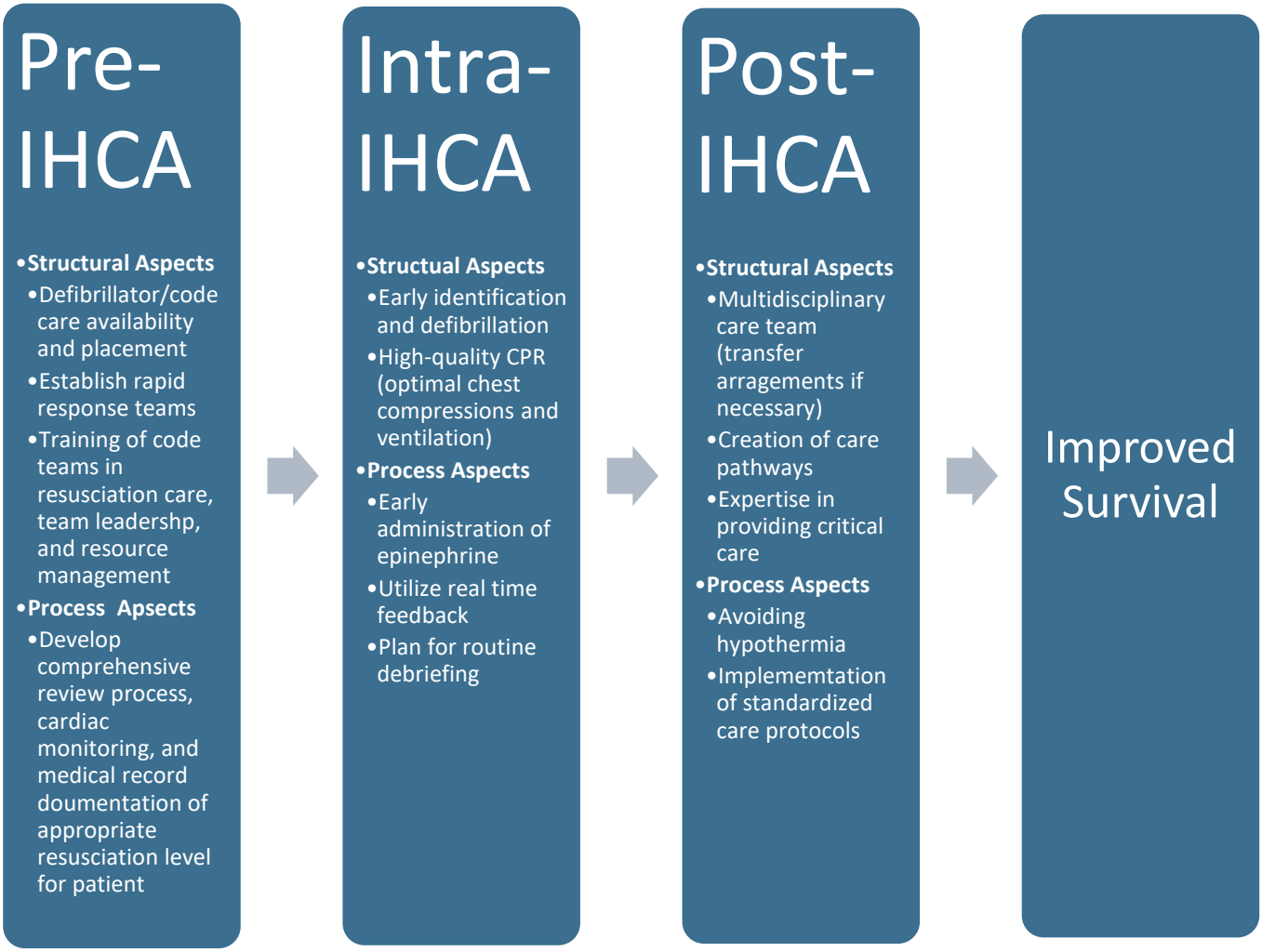
☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The diagram below shows how both structural and process aspects of care for all phases of in-hospital cardiac arrest (IHCA) can relate to improved survival.



Morrison LJ, Neumar RW, Zimmerman JL, Link MS, Newby LK, McMullan PW Jr, Vanden Hoek T, Halverson CC, Doering L, Peberdy MA, Edelson DP; on behalf of the American Heart Association Emergency Cardiovascular Care Committee, Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation Council on Clinical Cardiology, and Council on Peripheral Vascular Disease. Strategies for improving survival after in-hospital cardiac arrest in the United States: 2103 consensus recommendations: a consensus statement from the American Heart Association. *Circulation*. 2013;127:1538-1563.

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

Improvement in survival after in-hospital cardiac arrest can be affected by several structures and processes put in place by hospitals, such as the utilization of increased training of staff in resuscitation procedures (including the use of mock codes), earlier recognition of patients in cardiac arrest and shorter staff response time, and improved quality of chest compressions (Chan, 2015). Studies have shown that increased duration of resuscitation attempt, prompt administration of epinephrine, and timely delivery of defibrillation can all contribute to improve post-resuscitation rates. A study utilizing data from the Get With The Guidelines-Resuscitation Registry found that while an optimum duration of resuscitation attempt could not be determined, hospitals that had shorter resuscitation attempt duration (median 16 minutes) were less likely to see a return of spontaneous circulation compared to hospitals that had longer resuscitation attempt duration (median 25 minutes) (Goldberger, et. a., 2012). Regarding delayed administration of epinephrine, a

study found that these delays vary across hospitals, and that hospitals with higher rates of delay had worse survival rates (Khera R, Chan PS, Donnino M, Girota S, 2016). While it is widely known that prompt delivery of defibrillation contributes to improved survival rate, one study found that rates of delayed defibrillation vary across hospitals (Chan PS, Krumholz HM, Nichol G, Nallamothu BK, 2008; Chan PS, Nichol G, Krumholz HM, Spertus JA, Nallamotho BK, 2009). Additionally, survival rates post-in-hospital cardiac arrest have shown to improve with facility participation in the Get With The Guidelines-Resuscitation registry (from 16% up to 24% from 2010 to 2013) which could linked to improved resuscitation care (Girota, et. al., 2012).

Chan PS. Public health burden of in-hospital cardiac arrest. 2015. Available at: http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2015/GWTG.pdf

Chan PS, Krumholz HM, Nichol G, Nallamothu BK, and the American Heart Association National Registry of Cardiopulmonary Resuscitation Investigators. Delayed time to defibrillation after in-hospital cardiac arrest. N Engl J Med. 2008;358:9-17.

Chan PS, Nichol G, Krumholz HM, Spertus JA, Nallamouthu BK for the American Heart Association National Registry of Cardiopulmonary Resuscitation (NRCPR) Investigators. Hospital variation in time to defibrillation after in-hospital cardiac arrest. Arch Int Med. 2009;169:1265-1273.

Girota S, Nallomothu BK, Spertus JA, Li Y, Krumholz HM, Chan PS for the American Heart Association Get With The Guidelines—Resusciation Investigators. Trends in survival after in-hospital cardiac arrest. N Engl J Med. 2012 November 15;367(20):1912-1920.

Goldberger ZD, Chan PS, Berg RA, Kronick, SL, Cooke CR, Lu M, Bamerjee M, Hayward RA, Krumholz HM, Nallomouthou BK, for the American Heart Association Get With The Guidelines—Resusciation (formerly the National Registry of Cardiopulmonary Resusciation) Investigators. Duration of resuscitation efforts and survival after in-hospital cardiac arrest: an observational study. Lancet. 2012;380:1473-81.

Khera R, Chan PS, Donnino M, Girota S for the American Heart Association Get With The Guidelines-Resusciation Investigators. Hospital variation in time to epinephrine for nonshockable in-hospital cardiac arrest. Circulation. 2016;134:2105-2114.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

<u>If a COMPOSITE</u> (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

Survival rates after in-hospital cardiac arrest vary across hospitals and serve as not only and indicator of patient severity of illness, but also as an indicator of success for the resuscitation structures and processes a facility has in place. To date, there has not been a risk-standardized survival rate measure for this population by which facilities can compare themselves to others. This measure is intended to fill that gap.

Chan PS, Berg RA, Spertus JA, Schwamm LH, Bhatt DL, Fonarow GC, et. al. Risk standardizing survivial for in-hospital cardiac arrest to facilitate hospital comparisons. JACC. 2013. 62:601-609.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Based on the sample of 312 hospitals, the mean performance rate (adjusted survival) is 24%, the median performance rate is 24%. The standard deviation is 5%. The range of the performance rate is 27%, with a minimum rate of 11% and a maximum rate of 38%.

The range of performance from 11% to 38% suggests there is clinically meaningful variation across hospitals' performance.

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 maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Since the RSSR measure is a hospital-level measure, race-specific survival was not assessed at the patient-level. Instead, we divided hospitals between 2011 and 2015 with at least 10 IHCA patients into quartiles of patients of black race. The median hospital percentage of IHCA patients of black race was 11% (IQR: 4% to 27%). Hospitals with the smallest number of black patients (quartile 1) had a higher unadjusted (observed) and RSSR for IHCA as compared with hospitals that had the highest number of black patients (quartile 4), suggesting some degree of disparity in RSSRs by hospital racial composition (See Table 2 in the NQF Testing Attachment).

We therefore did not include race/ethnicity as a model covariate, because we did not want survival rates between hospitals to mask significant differences that may be due to race. In fact, if two hospitals do differ in their survival rates, race may be one reason why.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

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 sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

De.6. Non-Condition Specific*(check all the areas that apply):*

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

Not applicable.

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** RSSR_Specs_AHA_FINAL.pdf

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Patients who were alive at discharge

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Target population for the numerator is identified via the Get With The Guidelines (GWTG)—Resuscitation Registry using the time period and data fields below:

Time Period for Data Collection: At each hospital discharge during the measurement period.

'Discharge Status' = Alive

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

Patients aged 18 years and older with in-hospital cardiac arrest who received chest compression and/or defibrillation

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.) IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

Target population for the denominator is identified via the Get With The Guidelines (GWTG)—Resuscitation Registry using the time period and data fields below:

Time Period for Data Collection: 12 consecutive months

'Age at System Entry' > = 18 years

AND

'First documented pulseless rhythm' = Asystole, Pulseless Electrical Activity (PEA), Pulseless Ventricular Tachycardia, or Ventricular Fibrillation (VF)

AND

'Did patient receive chest compressions and/or defibrillation during this event?' = Yes

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

None

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

Not applicable.

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

Statistical risk model

If other:

S.12. Type of score:

Other (specify):

If other: Risk standardized rate

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

The measure score is calculated as follows:

1. Patients for inclusion are identified using inclusion criteria as described above (S.6 through S.9)

2. Patients meeting the numerator (S.4-S.5) are determined.

3. Variables for inclusion in risk adjustment are pulled.

4. Measure score is calculated using data aggregated from all registry participants, as described below and within the testing attachment.

The measure is adjusted using the variables below:

1. Age
2. Initial cardiac arrest rhythm
3. Hospital location
4. Hypotension
5. Sepsis
6. Metastatic or hematologic malignancy
7. Hepatic insufficiency
8. Mechanical ventilation
9. Intravenous vasopressor

Measure Calculation:

1) Create a model for predictors of in-hospital cardiac arrest (IHCA). Since patients at a given hospital with IHCA will have correlated outcomes, we use a multivariable hierarchical logistic regression model, wherein patients will be nested within hospitals in the model and hospitals are modeled as random effects.

2) A number of demographic (age category, sex) and comorbidity variables (includes pre-existing conditions and interventions in place at the time of cardiac arrest) are considered for model inclusion. Essentially, we consider almost all variables as potential predictors in the model.

3) An initial "full" model is generated with significant predictors of survival to discharge.

4) Within this initial "full" model, we then work to sequentially eliminate predictors with the smallest contribution to the model. This is done to derive a more parsimonious, or "reduced", model with 95% of the initial "full" model's predictive ability – in essence, to create a model with many fewer variables with almost identical predictive (discriminative) ability as the "full" model.

5) Model discrimination with the "reduced" model is then assessed with c-statistics, and model validation performed by comparing the R2 of the predicted and observed plots (this information is described in the next section).

6) Once the "reduced" predictive model is confirmed, as above, then one can calculate RSSRs for each hospital. This is accomplished by multiplying the weighted average unadjusted hospital survival rate for the entire study sample by the hospital's predicted vs. expected survival rate. So, a hospital with a predicted vs. expected survival rate > 1 would have a RSSR higher than the weighted mean, and one with a ratio < 1 would have a RSSR below the weighted mean.

7) The expected survival number (denominator) would be determined by applying the model's regression coefficients for covariates to each patient and summing up the probabilities for all patients within that hospital. This number uses the average hospital-level random intercept in the model.

8) The predicted survival number (numerator) is the number of survivors at a hospital, which is determined in the same way as the expected survival except that the hospital's specific random intercept is used.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

Not applicable.

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

Not applicable

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.

Registry Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*
IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

American Heart Association (AHA) Get With The Guidelines(R)-Resuscitation (GWTG-R) Registry

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

Available in attached appendix at A.1

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Emergency Department and Services, Inpatient/Hospital

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

Not Applicable

2. Validity – See attached Measure Testing Submission Form

0104_nqf_testing_attachment_7.1_RSSR.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:**
Measure Title: Risk-Standardized Survival Rate for In-Hospital Cardiac Arrest

Date of Submission: 11/6/2017

Type of Measure:

☒ **Outcome (*including PRO-PM*)**	☐ **Composite – *STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures**, section **2b3** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b5** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **instrument-based measures** (including PRO-PMs) **and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **instrument-based measures (including PRO-PMs) and composite performance measures**, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of **missing data** (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the sources of data specified and intended for*

measure implementation. ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: **(*must be consistent with data sources entered in S.17*)**	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ claims	☐ claims
☒ registry	☒ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Get With The Guidelines®-Resuscitation has its roots in the American Heart Association's National Registry of Cardiopulmonary Resuscitation (NRCPR), started in 1999 to collect resuscitation data from hospitals nationwide and create evidence-based guidelines for inpatient CPR.

In 2010, the program was incorporated into Get With The Guidelines and enhanced to provide additional resources, tools and benefits, including:

- identification of improvement opportunities
- performance comparison with hospitals
- reduction of noncompliance and medical errors through data-driven peer review
- access to the most up-to-date research and scientific publications
- professional education opportunities, such as workshops and webinars
- clinical tools and resources
- QI field staff support
- a competitive advantage in the healthcare marketplace
- national and local recognition for hospital team program achievement
- web-based data collection to fulfill Joint Commission standards and other requirements
- performance feedback reporting for continuous quality improvement

The Get With The Guidelines-Resuscitation program is provided by the American Heart Association/American Stroke Association.

1.3. What are the dates of the data used in testing? 01/2011 to 05/2015

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.20*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice

☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ 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*)

A total of 326 hospitals reported on this measure. Of these, 312 hospitals had all the required data elements and met the minimum number of quality reporting events (1) for inclusion the analyses. For this measure, 96 percent of hospitals were included in the analyses, and the average number of quality reporting events was 190, for a total of 61,934 cardiac arrest events, and 14,782 cardiac arrest survivals to discharge. Of the 312 hospitals, the range of cardiac arrest quality reporting events was 1 to 122, and the range for survivals was 0 to 344.

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*)

There were 48,841 cases included in this testing and analysis. These were the patients that were associated with hospitals who had 10 or more eligible cases for this measure.

Table 1 below describe the patient case mix (demographics and pre-existing conditions), stratified by the derivation and validation cohorts.

Table 1. **Population Clinical Characteristics**

*Patients with at least 10 events	**Derivation Cohort**	**Validation Cohort**
	(n = 32,560)	(n = 16,281)
Demographics		
Age, Mean ± SD	65.6 ± 16.1	65.2 ± 15.9
Male sex	18996 (58.3%)	36,241 (58.5%)
Race		
White	22576 (69.3%)	42,580 (68.8%)
Black	6678 (20.5%)	14,138 (22.8%)
Other	1268 (3.9%)	1530 (2.5%)
Unknown	2038 (6.3%)	3686 (6.0%)
Hispanic	2254 (6.9%)	2780 (4.5%)
Pre-Existing Conditions		
Respiratory insufficiency	13301 (40.9%)	26527 (42.8%)
Renal insufficiency	10850 (33.3%)	21336 (34.4%)
Diabetes mellitus	10001 (30.7%)	19652 (31.7%)
Hypotension	8413 (25.8%)	14645 (23.6%)
Heart failure this admission	5370 (16.5%)	9527 (15.4%)
Prior heart failure	6278 (19.3%)	12971 (20.9%)

*Patients with at least 10 events	Derivation Cohort	Validation Cohort
Myocardial infarction this admission	5184 (15.9%)	8807 (14.2%)
Prior Myocardial infarction	4791 (14.7%)	8389 (13.5%)
Metabolic or electrolyte abnormality	4765 (14.6%)	10640 (17.2%)
Septicemia	5519 (17.0%)	10550 (17.0%)
Pneumonia	4342 (13.3%)	8445 (13.6%)
Metastatic or hematologic malignancy	4046 (12.4%)	7108 (11.5%)
Hepatic insufficiency	2474 (7.6%)	4434 (7.2%)
Baseline depression in CNS function	3640 (11.2%)	5449 (8.8%)
Acute CNS non-stroke event	2250 (6.9%)	3797 (6.1%)
Acute stroke	1234 (3.8%)	2266 (3.7%)
Major trauma	1399 (4.3%)	2853 (4.6%)
Characteristics of arrest		
Cardiac arrest rhythm		
Asystole	10997 (33.8%)	17893 (28.9%)
Pulseless electrical activity	15327 (47.1%)	33240 (53.7%)
Ventricular fibrillation	3691 (11.3%)	6149 (9.9%)
Pulseless ventricular tachycardia	2545 (7.8%)	4652 (7.5%)
Location		
Intensive care unit	15780 (48.5%)	30084 (48.6%)
Monitored unit	5034 (15.5%)	9442 (15.2%)
Non-Monitored unit	5632 (17.3%)	9477 (15.3%)
Emergency room	3307 (10.2%)	7072 (11.4%)
Procedural or surgical area	2132 (6.5%)	4662 (7.5%)
Other	675 (2.1%)	1197 (1.9%)
Interventions in Place		
Mechanical ventilation	10747 (33.0%)	20604 (33.3%)
IV Vasopressor	9549 (29.3%)	14177 (22.9%)
Dialysis	1163 (3.6%)	1687 (2.7%)

*2011-2015 registry data.

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 same data sample was used for all testing.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

Since the RSSR measure is a hospital-level measure, race-specific survival was not assessed at the patient-level. Instead, we divided hospitals between 2011 and 2015 with at least 10 IHCA patients into quartiles of patients of black race. The median hospital percentage of IHCA patients of black race was 11% (IQR: 4% to 27%). Hospitals with the smallest number of black patients (quartile 1) had a higher unadjusted (observed) and RSSR for IHCA as compared with hospitals that had the highest number of black patients (quartile 4), suggesting some degree of disparity in RSSRs by hospital racial composition (see table below).

We therefore did not include race/ethnicity as a model covariate, because we did not want survival rates between hospitals to mask significant differences that may be due to race. In fact, if two hospitals do differ in their survival rates, race may be one reason why.

Table 2.

	Hospital Quartile of % of Black IHCA Patients				All Hospitals n = 288	P
	Least Black			Most Black		
	Q1; n=72	Q2; n = 72	Q3; n = 72	Q4; n = 72		
Observed Rate						< 0.001
Mean ± SD	26% ± 9%	24% ± 8%	24% ± 7%	20% ± 7%	24% ± 8%	
Median (IQR)	27% (20%, 31%)	23% (19%, 28%)	24% (20%, 28%)	20% (17%, 23%)	23% (19%, 28%)	
RSSR						0.002
Mean ± SD	25% ± 5%	24% ± 5%	25% ± 6%	22% ± 5%	24% ± 5%	
Median (IQR)	25% (22%, 29%)	24% (20%, 28%)	25% (21%, 29%)	23% (19%, 26%)	24% (21%, 28%)	

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 hospital performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in hospital performance.

Our signal-to-noise ratio testing was conducted by fitting a hierarchical, logistic regression model to derive the two shape parameters – alpha and beta; the model was built on a specified beta-binomial distribution. This modeling was done on patient-level data, adjusting for age, gender, location of arrhythmia (i.e. ICU, ED), type of heart rhythm, and present on arrival (POA) conditions. This methodological approach enabled us to account for patient-level and hospital-level (random effects) mixed effects.

The two estimated model parameters were then used to calculate between-site (hospital-to-hospital) and within-site (hospital-specific) variances. The formulas used are described below:

Reliability = (hospital-to-hospital variance) /((hospital-to-hospital variance) + (hospital-specific variance))

Between-site (hospital-to-hospital) variance = $\alpha\beta/(\alpha + \beta + 1)(\alpha+\beta)^2$

Within-site or (hospital-specific) variance = $\hat{p}_i(1 - \hat{p}_i)/n_{is}$

Where,

p-hat$_i$ = the proportion of patients who survived to discharge at hospital $_i$.

n_{is} = the total number of cardiac arrest events at hospital $_i$.

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*)

This measure had a signal-to-noise ratio reliability of 0.70, at the average number of events. At the minimum number of events [10], reliability was 0.693.

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 at the average number of quality events is moderate.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

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

The expert panel included 34 members. Panel members were comprised of experts from the PCPI Cardiovascular Technical Expert Panel and the AHA Emergency Cardiac Care Committee.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Frequency Distribution of Ratings

1 – 1 responses (Strongly Disagree)

2 – 3 responses (Disagree)

3 – 6 responses (Neither Agree nor Disagree)

4 – 17 responses (Agree)

5 – 7 responses (Strongly Agree)

The results of the expert panel rating of the validity statement were as follows: N = 34; Mean rating = 6.8 and 71% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

We also provide a table below of the respondents' specialty. Our face validity survey was administered to a diverse group of experts, reducing bias.

Table 3

Specialty of Respondents	Count
Pharmacy	1
Psychology	1
Pulmonary medicine	1
Preventive medicine	1
Nursing	2
Research science/outcomes	2
Anesthesia	3
Internal/family medicine	4
Cardiology (include pediatrics)	7
Emergency medicine	12
Total	34

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Based on the mean rating by the expert panel, this measure is valid as specified.

_____

2b2. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section* 2b3

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

No exclusions.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

No exclusions.

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

No exclusions.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with _risk factors**

☐ **Stratification by _risk categories**

☒ **Other,** Standardized

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

There are several steps taken to derive and calculate a hospital's risk standardized survival rate (RSSR) for in-hospital cardiac arrest (IHCA). Those steps are detailed below:

1) Create a model for predictors of IHCA. Since patients at a given hospital with IHCA will have correlated outcomes, we use a multivariable hierarchical logistic regression model, wherein patients will be nested within hospitals in the model and hospitals are modeled as random effects.
2) A number of demographic (age category, sex) and comorbidity variables (includes pre-existing conditions and interventions in place at the time of cardiac arrest) are considered for model inclusion. Essentially, we consider almost all variables (except race) as potential predictors in the model. The table below lists all covariates (also termed predictors or risk factors) included in the full model, along with their corresponding estimates (also termed coefficients), their odds ratios (amount of risk relative to the reference population), and their 95% confidence intervals. See table 4 below.
3) An initial "full" model is generated with significant predictors of survival to discharge.
4) Within this initial "full" model, we then work to sequentially eliminate predictors with the smallest contribution to the model. This is done to derive a more parsimonious, or "reduced", model with 95% of the initial "full" model's predictive ability – in essence, to create a model with many fewer variables with almost identical predictive (discriminative) ability as the "full" model.
5) Model discrimination with the "reduced" model is then assessed with c-statistics, and model validation performed by comparing the R^2 of the predicted and observed plots (this information is described in the next section).
6) Once the "reduced" predictive model is confirmed, as above, then one can calculate RSSRs for each hospital. This is accomplished by multiplying the weighted average unadjusted hospital survival rate for the entire study sample by the hospital's predicted vs. expected survival rate. So, a hospital with a predicted vs. expected survival rate > 1 would have a RSSR higher than the weighted mean, and one with a ratio < 1 would have a RSSR below the weighted mean.
7) The expected survival number (denominator) would be determined by applying the model's regression coefficients for covariates to each patient and summing up the probabilities for all patients within that hospital. This number uses the average hospital-level random intercept in the model.
8) The predicted survival number (numerator) is the number of survivors at a hospital, which is determined in the same way as the expected survival except that the hospital's specific random intercept is used.

Table 4

Predictor	Estimate	Odds Ratio	95% CI
Age			
<50	0	Reference	Reference

Predictor	Estimate	Odds Ratio	95% CI
50-59	-0.0202	0.98	0.88-1.08
60-69	-0.0408	0.96	0.87-1.05
70-79	-0.2877	0.75	0.68-0.83
≥ 80	-0.6931	0.5	0.46-0.56
Male sex	-0.0834	0.92	0.87-0.98
Hospital location			
Non-monitored unit	0	Reference	Reference
Intensive care unit	0.5653	1.76	1.59-1.93
Monitored unit	0.47	1.6	1.45-1.78
Emergency room	0.5188	1.68	1.49-1.89
Procedural or surgical area	1.1217	3.07	2.71-3.49
Other	0.6259	1.87	1.54-2.26
Initial cardiac arrest rhythm			
Asystole	0	Reference	Reference
Pulseless electrical activity	0.0392	1.04	0.97-1.12
Ventricular fibrillation	1.2238	3.4	3.10-3.72
Pulseless ventricular tachycardia	1.1086	3.03	2.73-3.36
Myocardial infarction this admission	0.1484	1.16	1.07-1.25
Prior heart failure	-0.0619	0.94	0.87-1.01
Renal insufficiency	-0.2231	0.8	0.75-0.86
Hepatic insufficiency	-0.6539	0.52	0.45-0.59
Hypotension	-0.4463	0.64	0.59-0.69
Septicemia	-0.4308	0.65	0.59-0.71
Acute stroke	-0.3147	0.73	0.63-0.86
Diabetes mellitus	0.131	1.14	1.06-1.21
Metabolic/electrolyte abnormality	-0.1625	0.85	0.77-0.94
Metastatic or hematologic malignancy	-0.755	0.47	0.42-0.53
Major trauma	-0.3425	0.71	0.60-0.83
Mechanical ventilation	-0.5447	0.58	0.54-0.63
Dialysis	-0.3011	0.74	0.61-0.90
IV Vasopressor	-0.734	0.48	0.44-0.52

All variables are defined by the AHA-GWTG registry, in particular, the Resuscitation Patient Management Tool ®. Reporting hospitals would complete the CPA event form. This data would then be entered into the registry.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

Clinical and statistical experts (from disciplines such as cardiology, neurology, critical care, and research) selected model covariates based on patient clinical characteristics that most influence survival during an in-hospital cardiac arrest. These patient factors can be categorized by the following:

1) Patient demographics (i.e. age, gender)
2) Location of the cardiac arrest (i.e. intensive care, ED)
3) Initial cardiac rhythm (i.e. VT, VF)
4) Pre-existing conditions/present on arrival (POA) conditions (i.e. heart failure, sepsis)
5) Critical-care interventions in place prior to the arrest (i.e. mechanical ventilation, intravenous vasopressor support, pulmonary artery catheter)

All of these factors were carefully considered from both a clinical perspective and a statistical perspective. Careful thought went into ensuring all significant risk factors were included, but that the model would not be overfit.

The risk factors mentioned above were included in the initial, full model. Model reduction involved a process of keeping only significantly contributing risk factors in the final model. This was done to derive a more parsimonious, or "reduced", model with 95% of the initial "full" model's predictive ability – in essence, to create a model with fewer variables with almost identical predictive (discriminative) ability as the "full" model. The purpose of this process was to derive retain as much predictive ability, without overfitting the model.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☒ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

Table 5 below lists the risk factors that were included in the final model, along with their estimates, ORs, and 95% CIs. For our model reduction methodology, please see the final paragraph of section 2b4.3 above.

Table 5

Predictor	Beta-Weight Estimate	Odds Ratio	95% CI
Age			
<50	0	Reference	Reference
50-59	0.0031	1	0.91-1.11
60-69	-0.0096	0.99	0.90-1.09
70-79	-0.256	0.77	0.70-0.85

Predictor	Beta-Weight Estimate	Odds Ratio	95% CI
≥ 80	-0.6562	0.52	0.47-0.57
Initial cardiac arrest rhythm			
Asystole	0	Reference	Reference
Pulseless electrical activity	0.0478	1.05	0.98-1.13
Ventricular fibrillation	1.2631	3.54	3.24-3.86
Pulseless ventricular tachycardia	1.1289	3.09	2.79-3.43
Hospital location			
Non-monitored unit	0	Reference	Reference
Intensive care unit	0.5643	1.76	1.60-1.93
Monitored unit	0.4816	1.62	1.46-1.79
Emergency room	0.5618	1.75	1.56-1.97
Procedural or surgical area	1.155	3.17	2.80-3.60
Other	0.621	1.86	1.54-2.25
Hypotension	-0.4749	0.62	0.57-0.67
Sepsis	-0.4879	0.61	0.56-0.68
Metastatic or hematologic malignancy	-0.7345	0.48	0.43-0.53
Hepatic insufficiency	-0.724	0.48	0.42-0.56
Mechanical ventilation	-0.5662	0.57	0.53-0.61
IV Vasopressor	-0.7329	0.48	0.44-0.52

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Based on the information provided in 1.8, the decision was made to not adjust the measure based on SDS factors, as identification of differences on these factors is an important indicator of identifying variability in quality.

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Model discrimination was assessed with the C-statistic, and model validation was performed in the remaining one-third of the study cohort by examining observed vs. predicted plots.

Of 48,841 patients in the study cohort, 32,560 were randomly selected for the derivation cohort and 16,281 for the validation cohort.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Initially, 18 independent predictors were identified in the derivation cohort with the multivariable model, resulting in a model C-statistic of 0.708. After model reduction to generate a parsimonious model with no more than 5% loss in model prediction power, our final model was comprised of 9 variables, with only a small change in the C-statistic (0.704). When the model was tested in the independent validation cohort, model discrimination was similar (C-statistic of 0.707).

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Below describe our risk model calibration statistics. We describe 2012, 2013, and 2014 separately, and 2011-2015 as a whole.

2012 DATA

We re-developed a parsimonious model (c-statistic 0.694). The model for 2012 data calibrated well, with an R^2 of 0.99 (below). The discrimination and validation analyses using 2012 data prospectively validates the initial RSSR model using data between 2007 and 2010.

2013 DATA

We re-developed a parsimonious model (c-statistic 0.709). The model for 2013 data also calibrated well, with an R^2 of 0.99 (below). The discrimination and validation analyses using 2013 data prospectively validates the prior RSSR model using data between 2007 and 2010.

2014 DATA

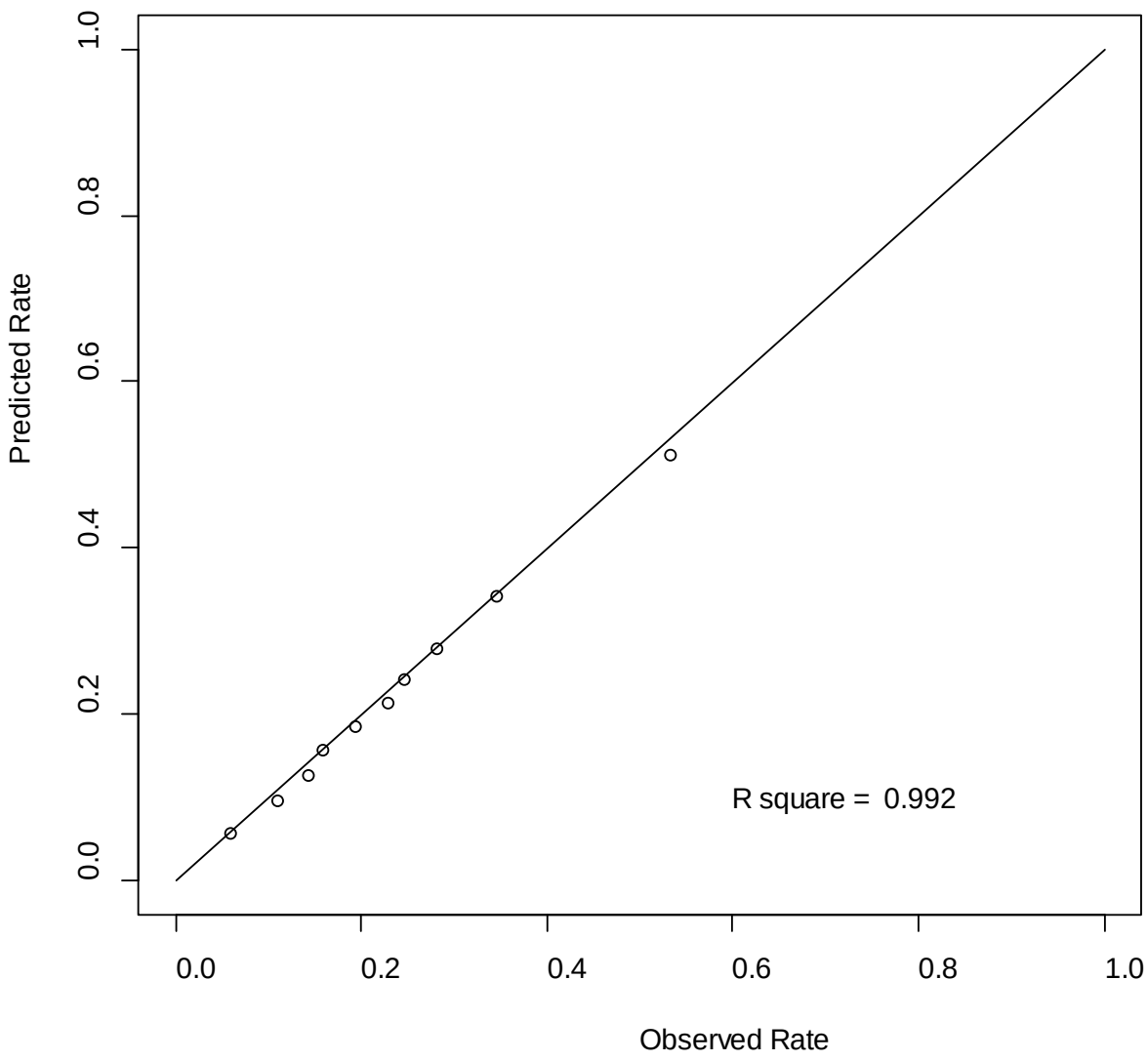
We re-developed a parsimonious model using 2014 data (c-statistic 0.703). The model for 2014 data also calibrated well, with an R^2 of 0.99 (below). The discrimination and validation analyses using 2014 data prospectively validates the prior RSSR model using data between 2007 and 2010.

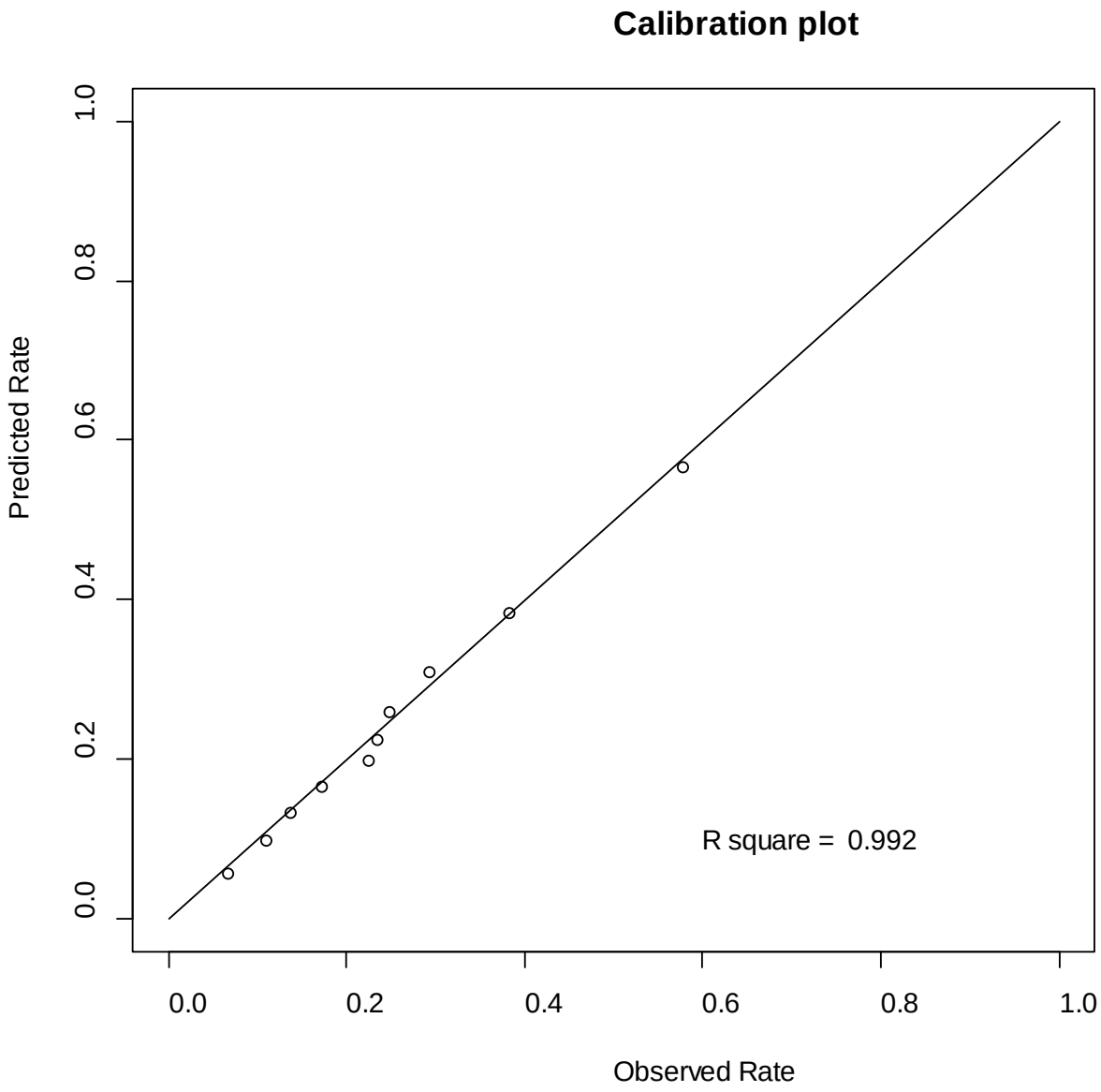
2011-2015

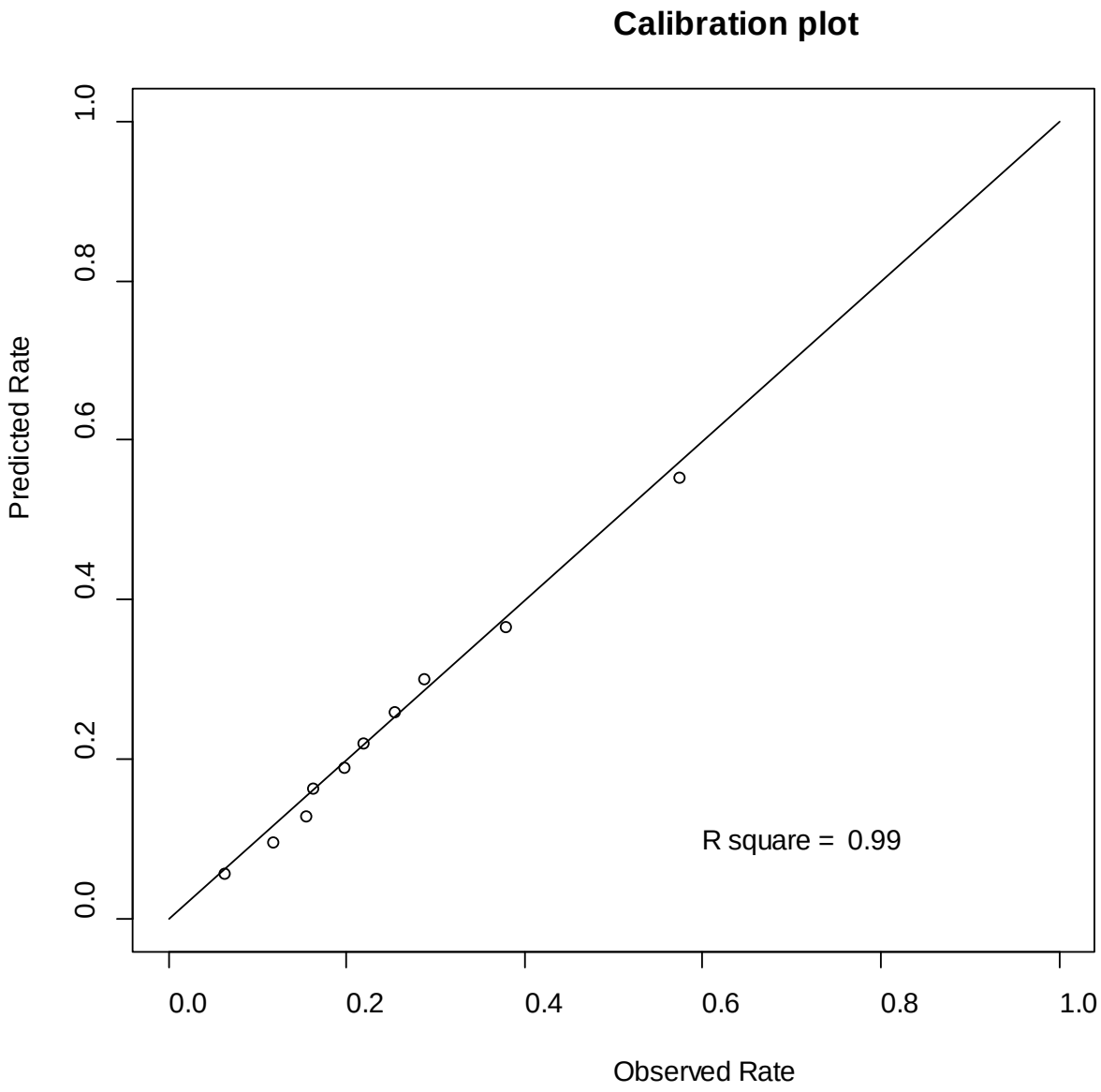
We re-developed a parsimonious model using the data from 2011 to 2015 (c-statistic 0.706). The model using 2011-2015 data also calibrated well, with an R^2 of 0.997 (below). The discrimination and validation analyses using combined 2011-2015 data prospectively validates the initial RSSR model, which used data between 2007 and 2010.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2012 DATA

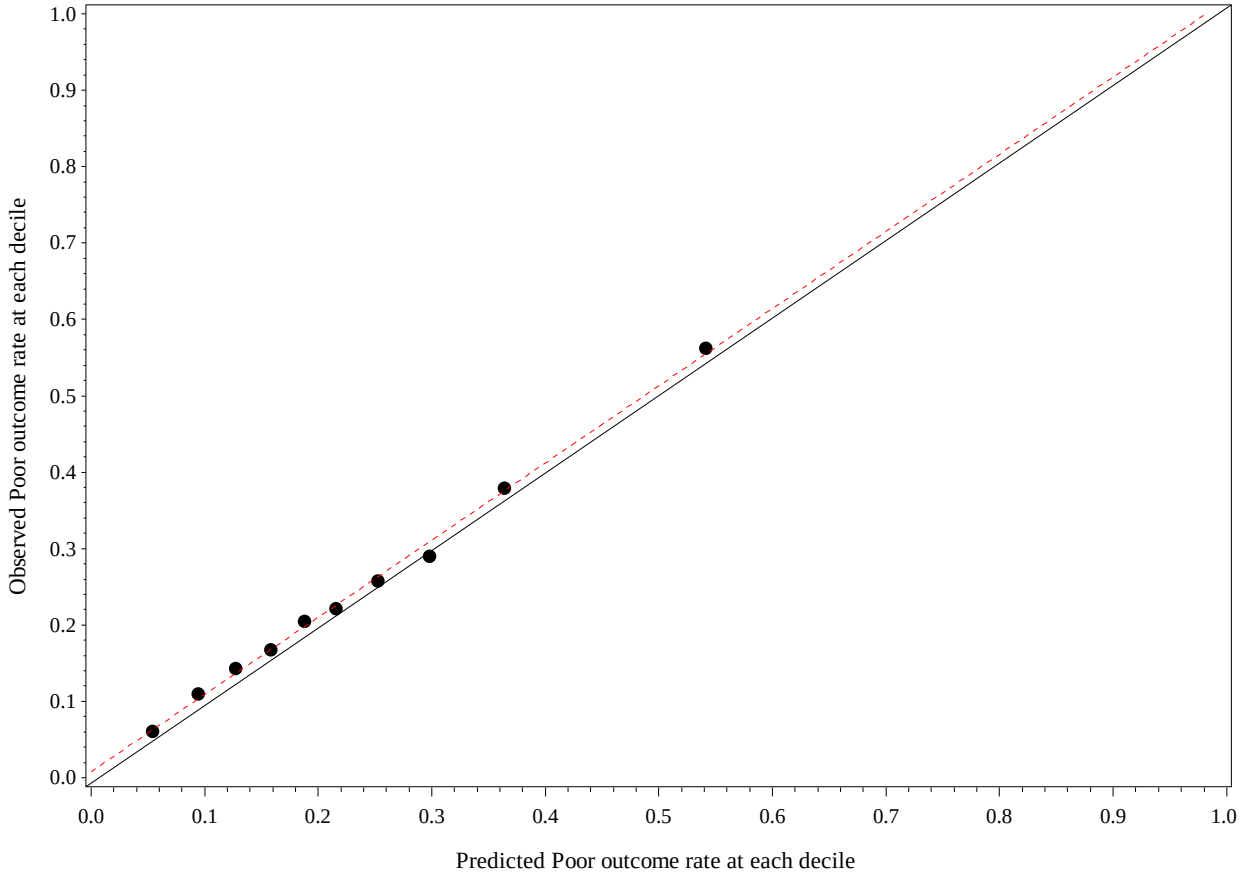

Calibration plot
Predicted Rate
0.0
0.2
0.4
0.6
0.8
1.0
R square = 0.992
Observed Rate
0.0
0.2
0.4
0.6
0.8
1.0


Calibration plot
Predicted Rate
0.0
0.2
0.4
0.6
0.8
1.0
R square = 0.992
Observed Rate


Calibration plot
Predicted Rate
0.0
0.2
0.4
0.6
0.8
1.0
R square = 0.99
Observed Rate

2011-2015

Model Reduced: Survival to Discharge (deciles are determined by predicted probabilities)



2b3.9. Results of Risk Stratification Analysis:

Models were not risk-stratified.

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

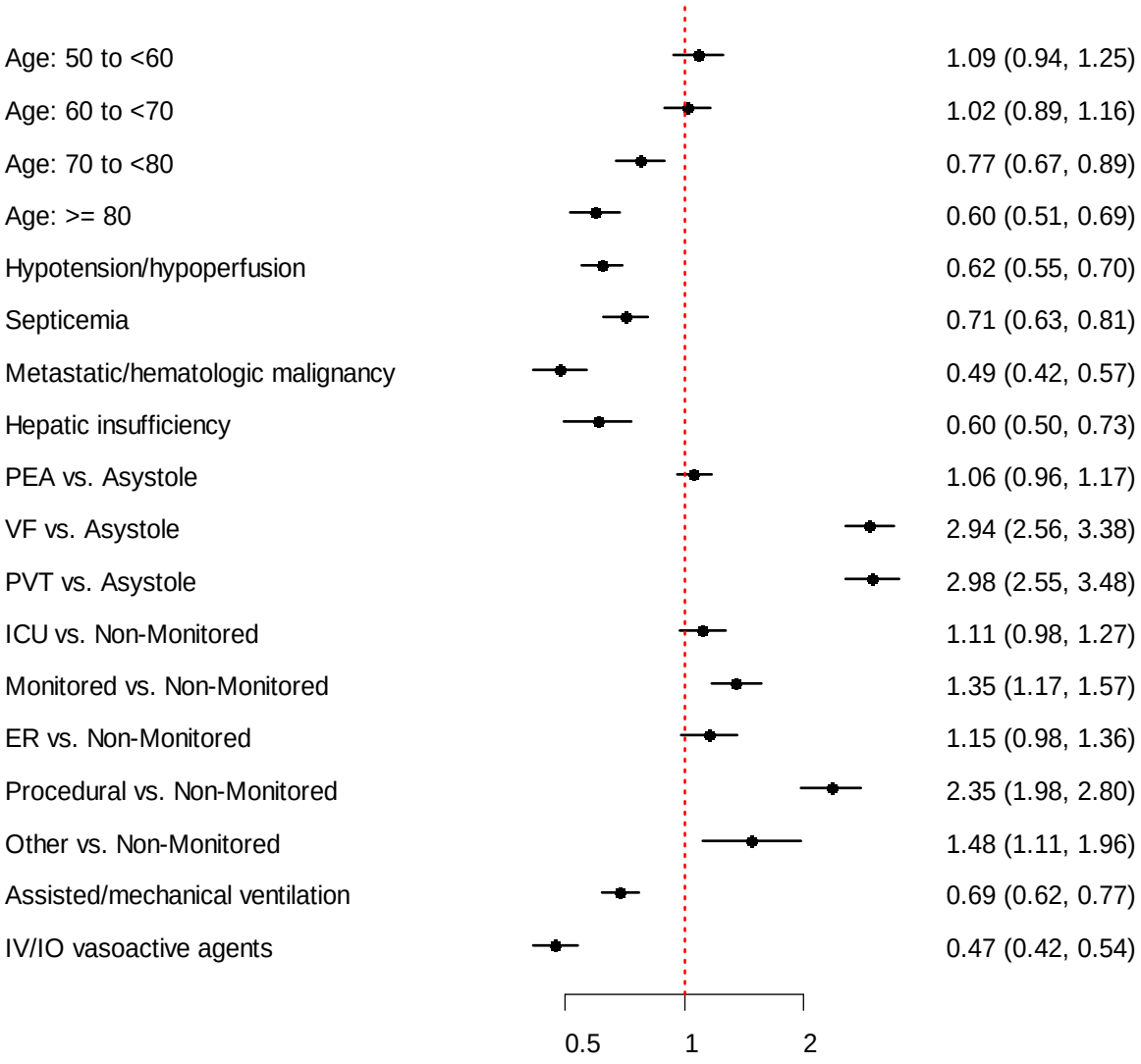
The results above indicate that the risk models are valid, predictive, descriptive, and are well-calibrated.

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

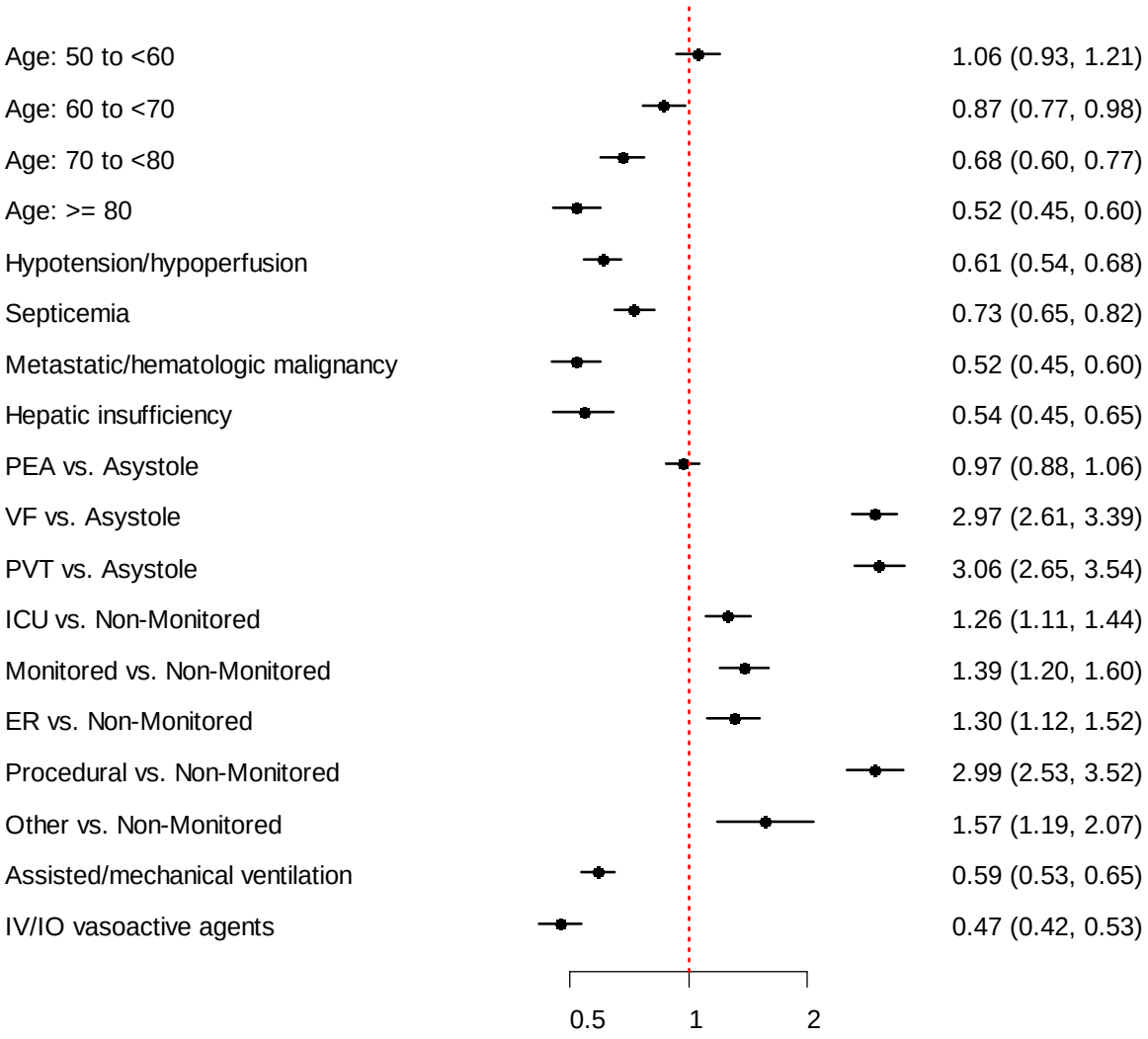
Forest plots

We provide below forest plots for the following years: 2012, 2013, and 2014. These plots illustrate the statistical significance of each variable, compared to its reference.

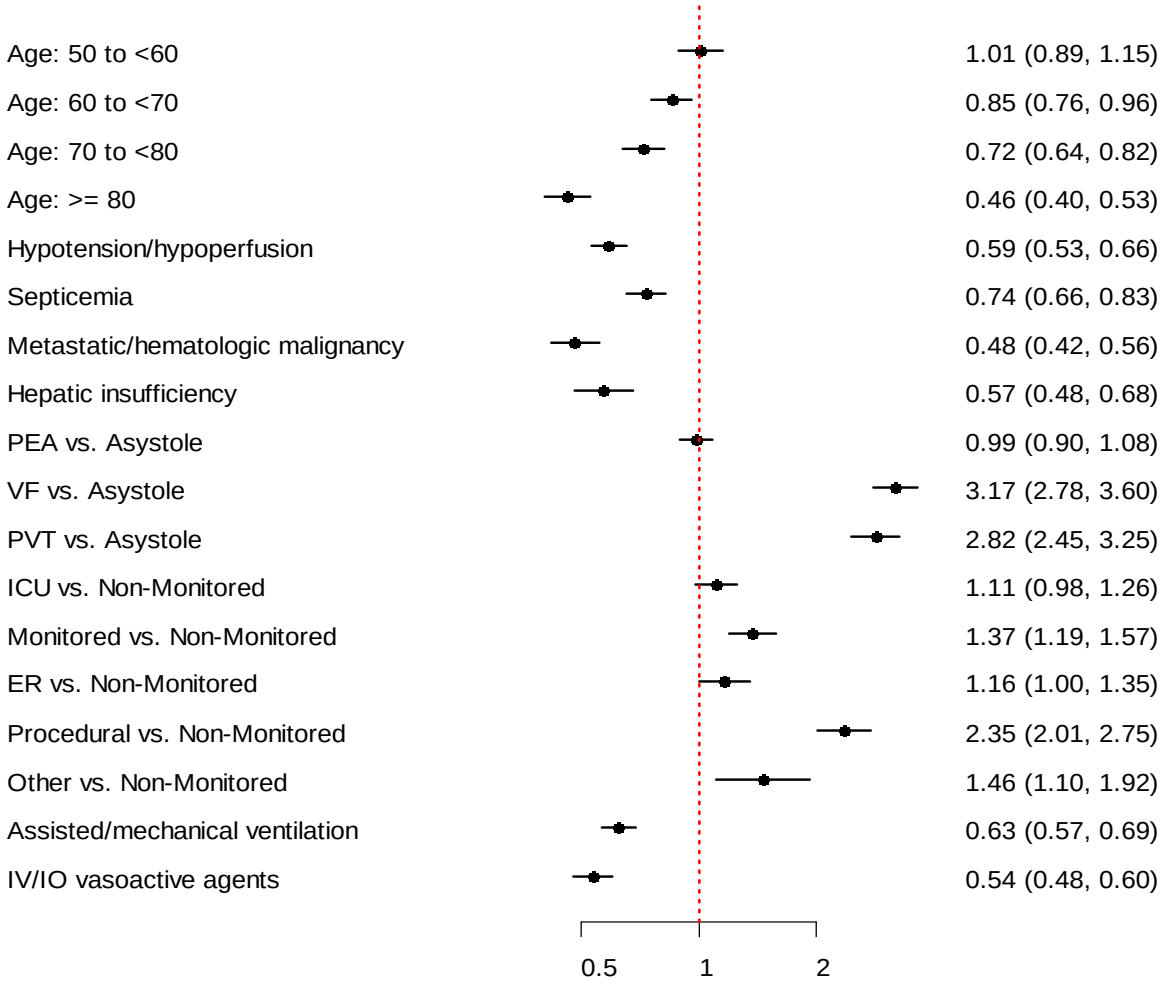
2012 Forest Plot



2013 Forest Plot



2014 Forest Plot



2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Measures of central tendency, variability, and dispersion were calculated (see table 6 below).

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Based on the sample of 312 hospitals, the mean performance rate (adjusted survival) is 24%, the median performance rate is 24%. The standard deviation is 5%. The range of the performance rate is 27%, with a minimum rate of 11% and a maximum rate of 38%.

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The range of performance from 11% to 38% suggests there is clinically meaningful variation across hospitals' performance.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

N/A.

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A.

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A.

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Data are not available to complete this testing.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Data are not available to complete this testing.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

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), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

Given that the data for this measure is collected through the Get With the Guidelines – Resuscitation Registry, and is not collected in an electronic health record, no feasibility assessment was performed. No issues with data collection have been identified and no modifications have been made to this measure, as collected in the GWTG – Resuscitation Registry, due to issues with data collection, sampling or cost.

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.

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Quality Improvement (Internal to the specific organization)	Professional Certification or Recognition Program American Heart Association Get With The Guidelines-Resuscitation http://www.heart.org/HEARTORG/Professional/GetWithTheGuidelines-Resuscitation/Get-With-The-Guidelines-Resuscitation_UCM_314496_SubHomePage.jsp Quality Improvement (external benchmarking to organizations) American Heart Association Get With The Guidelines-Resuscitation http://www.heart.org/HEARTORG/Professional/GetWithTheGuidelines-Resuscitation/Get-With-The-Guidelines-Resuscitation_UCM_314496_SubHomePage.jsp

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

Name of the program and sponsor: American Heart Association Get With The Guidelines-Resuscitation Registry
Purpose: Get With The Guidelines®-Resuscitation is the American Heart Association's collaborative quality improvement program demonstrated to improve adherence to evidence-based care of patients who experience an in-hospital resuscitation event or received post cardiac arrest care following an in-hospital or out-of-hospital event. The program facilitates the efficient capture, analysis and reporting of data that empowers and
supports the implementation of current guidelines, creation and dissemination of new knowledge, and development of next generation, evidence-based practice in resuscitation science. Hospitals are able to track data for Cardiopulmonary Arrest (CPA), Medical Emergency Team (MET), Post-Cardiac Arrest Care (PCAC) and Acute Respiratory Compromise (ARC) in the Web-based Patient Management Tool™ (powered by Quintiles Real-World & Late Phase Research). The PMT provides decision support, robust registry, real-time benchmarking capabilities and other performance improvement methodologies toward the goal of enhancing patient outcomes and saving lives.
The primary goal of Get With The Guidelines-Resuscitation is to save more lives by preventing in-hospital cardiac arrest and optimizing outcomes through benchmarking, quality improvement, knowledge translation, and research.
Level of measurement: Hospital (facility). There are currently 373 hospitals participating in the registry that are geographically diverse.

Name of Program and Sponsor: Recognition Program: American Heart Association Get With The Guidelines-Resuscitation Recognition Program
Purpose: Hospitals that participate actively and consistently in Get With The Guidelines®--Resuscitation are eligible for public recognition. Participating in GWTG-R is the first level of recognition. It acknowledges program participation and entry of baseline data into the Patient Management ToolTM. This recognition program launched on January 1, 2016. Awards recognize hospitals that demonstrate at least 85 percent compliance in each of the four Get With The Guidelines-Resuscitation Recognition Measures. The different levels reflect the amount of time for which the hospital demonstrates
performance.

- Bronze recognizes performance of 1 calendar quarter.
- Silver recognizes performance of 1 calendar year (January 1st to December 31st).
- Gold recognizes performance of 2 consecutive calendar years (January 1st to December 31st).

In 2017, 128 participating hospitals received public recognition in the program; 11 Bronze, 66 Silver, and 51 Gold.
Recognition Measures include:
Adult or Pediatric

- CPA: Time to first chest compressions <= 1 min in adult or pediatric patients and newborn/neonates >= 10 min old: Percent of events in adult or pediatric patients where time to first chest compressions <= 1 minute of event recognition.
- CPA: Device confirmation of correct endotracheal tube placement: Percent of adult or pediatric events with an endotracheal tube placement which was confirmed to be correct.
- CPA: Time to first shock <= 2 min for VF/pulseless VT first documented rhythm: Percent of events in adult or pediatric patients with VF/pulseless VT first documented rhythm in whom time to first shock <= 2 minutes of event recognition.
- CPA: Percent pulseless cardiac events monitored or witnessed: Percent of events in adult or Pediatric patients who were monitored or witnessed at the time of arrest.
- Newborn/Neonate
- CPA: Time to first chest compressions <= 1 min in adult and pediatric patients and newborn/neonates >= 10 min old: Percent of events in newborn/neonates >= 10 minutes old where time to first chest compressions <= 1 minute of event recognition.
- CPA: Time to first chest compressions <= 2 min for newborn/neonates < 10 min old: Percent of events in newborn/ neonates < 10 minutes old with time to first chest compressions <= 2 minutes of event recognition.
- CPA: Time to invasive airway <= 2 min in newborn/neonates from onset of cardiac event: Percent of newborn/neonatal events with an invasive airway inserted within 2 minutes of event recognition.
- CPA: Device confirmation of correct endotracheal tube placement: Percent of events with an endotracheal tube placement which was confirmed to be correct.

Please note: Recognition criteria are subject to change based on program enhancements.
Level of Measurement: Hospital (facility).

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not applicable.

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

The American Heart Association is currently in the early planning stages of a voluntary public reporting program for the Get With The Guidelines-Resuscitation program and is additionally planning on adding this measure to the Get With The Guidelines-Resuscitation recognition program in the near future.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

The measure and its specifications and results were vetted with the American Heart Association Research Committee chairs and feedback was provided. This feedback was incorporated into the final measure

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

Participants in the Get With The Guidelines-Resuscitation program have access to their data through the registry (also called the Patient Management Tool), where they are able to query and review results. Additionally, they receive a separate feedback report, available as a pdf download, of their risk-standardized in-hospital cardiac arrest results (example attached).

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Not applicable.

4a2.2.2. Summarize the feedback obtained from those being measured.

Not applicable.

4a2.2.3. Summarize the feedback obtained from other users

Not applicable.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

Not applicable.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Survival rates after in-hospital cardiac arrest had started to improve prior to the introduction of the feedback reports regarding results on the risk-standardized in-hospital cardiac arrest survival. Nonetheless, the wide variation in results underscores the importance of this measure and the feedback of its results to facilities in order to support efforts to improve patient survival rates after in-hospital cardiac arrest.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

Not applicable.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

Not applicable.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

Not applicable.

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.

Attachment Attachment: AHA_RSSR_Supplemental_Appendix.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American Heart Association

Co.2 Point of Contact: Melanie, Shahriary, melanie.shahriary@heart.org, 301-651-7548-

Co.3 Measure Developer if different from Measure Steward: American Heart Association

Co.4 Point of Contact: Melanie, Shahriary, melanie.shahriary@heart.org, 301-651-7548-

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.

Development of this measure by the American Heart Association GWTG-Resuscitation Investigators led to publication of the methodology (article attached).
Representing the American Heart Association GWTG-Resuscitation Investigators (Get With The Guidelines-Resuscitation Adult Task Force)
Paul S. Chan, MD, MS
Robert A. Berg, MD
John A. Spertus, MD, MPH
Lee H. Schwamm, MD
Deepak L. Bhatt, MD, MPH
Gregg C. Fonarow, MD
Paul A. Heidenreich, MD, MS
Brahamajee K. Nallomothu, MD, MPH
Fengming Tang, MS
Raina M. Merchant, MD, MSHP
Comilla Sasson MD, MS
Steven Bradley, MD, MPH
Michael W. Donnino, MD
Dana P. Edelson MD, MS
Robert T. Faillace MD, ScM
Romergryko Geocadin, MD

AHA Staff: Tanya Truitt

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2013

Ad.3 Month and Year of most recent revision: 2017

Ad.4 What is your frequency for review/update of this measure? Annual Review

Ad.5 When is the next scheduled review/update for this measure? 2018

Ad.6 Copyright statement: © 2017 American Heart Association/American Stroke Association. All Rights Reserved.

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: