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 an 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).
- In the fall 2017 cycle, the Committee initially reviewed this measure before it was withdrawn from consideration by the developer. At that time, the Committee agreed that there was evidence to support this measure despite the limitations outlined by the literature and developer.

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

RATIONALE:

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 326 hospitals from 2011-May 2015, 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%

- Hospitals included in the analysis varied in teaching status, bed size, level of trauma center, and region. Most hospitals (approximately 92%) were located in urban areas.
- The developer provides demographics and clinical characteristics, including pre-existing conditions, characteristics of arrest, and interventions in place for patients in the prospective validation cohort (n=61,934).

Disparities

- Race-specific survival was not assessed at the patient-level. The developer divided hospitals between 2011 and 2015 with at least 20 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?
- Is the developer's approach to addressing and identifying disparities reasonable?

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:

- i couldn't find the actual evidence - presumably it exists
- Moderate
- There is direct evidence that this measure is related to a desirable outcome that is valued by the target population.
- Outcomes measure with relationship between the measured health outcome and at least one healthcare action demonstrated.
- specific processes support the outcome
- Evidence is directly related to the measure and provider can change processes to improve outcoomes

1b. Performance Gap:

- a wide performance gap was shown across different time windows
- Moderate
- There is a significant performance gap. Optimally, but not realistically mortality would be 0%
- Significant performance gap and opportunity for improvement. Some disparity implied from data, but not exclusively examined.
- Gap in care demonstrated (11%-38%). Not stratified by race, but the performance of hospitals with greater numbers of patienits of black race performed worse.
- Performance ranges from 11% resuscitated to 38%. This represents a considerable range. 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.

Criteria 2: Scientific Acceptability of Measure Properties

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

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

Reliability

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

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

Validity

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

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

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

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

- Reliability testing was conducted at the measure score level using a signal-to-noise (SNR) analysis (specifically, Adams' beta-binomial method)
    - Testing data
        - 2011-2015 analysis:
            - 326 hospitals included; total of 61,934 cardiac arrest events and 14,782 (23.9%) patients survived to hospital discharge
            - Average number of quality reporting events per hospital =190
            - Range of quality reporting events was 1 to 1222
            - Range for number of patients surviving to hospital discharge was 0 to 344.
        - 2013 analysis
            - 273 hospitals included; total of 17,992 cardiac arrest events and 4417 (24.5%) patients survived to hospital discharge
            - Average number of quality reporting events per hospital =66
            - Range of quality reporting events was 1 to 360
            - Range for number of patients surviving to hospital discharge was 0 to 121
            - Analysis restricted the analyses to the 206 hospitals that had a minimum number of 20 quality reporting events
        - 2014 analysis
            - 259 hospitals included; total of 17,244 cardiac arrest events and 4163 (24.1%) patients survived to hospital discharge

- Average number of quality reporting events per hospital =67
- Range of quality reporting events was 1 to 409
- Range for number of patients surviving to hospital discharge was 0 to 124
- Analysis restricted the analyses to the 200 hospitals that had a minimum number of 20 quality reporting events

- Results:
  - Using entire prospective validation period (2011-2015): SNR mean= 0.76; median= 0.78
  - 2013: SNR mean = 0.70; median = 0.72
  - 2014: SNR mean = 0.67; median = 0.68
- Panelists expressed desire for more than means and medians (e.g., distribution of reliability estimates).

<u>Validity</u>

- The developer assessed the face validity of the measure score.
  - The face validity assessment adheres to NQF requirements
  - 34 of 34 TEP members responded
  - Mean rating = 3.76 (out of 5)
  - 71% of respondents (n=24) either agreed or strongly agreed 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*".
  - 18% of respondents (n=6) neither agreed or disagreed
    - Some of these respondents had no expertise in risk-adjustment/clinical matters; one was concerned that measure doesn't account for DNR rates across hospitals
  - 12% of respondents (n=4) disagreed or strongly disagreed
    - Concerns with lack of adjustment for DNR or other risk factors (e.g., race)
- Risk adjustment
  - Risk adjusted with 9 risk factors
    - Developers did not include social risk factors in the adjustment approach.
      - Sub-group analysis underscores disparities in survival based on race, but the developer notes the measure should not provide a rationale for accepting existing disparities in care.
      - The developer notes clinicians responding to in-hospital cardiac arrest would not be aware of a patient's social or economic risk and that additional variables (e.g., income, employment, education) are not collected.
  - Discrimination statistics (c-statistics)
    - Initial sample: 0.704
    - 2012: 0.694
    - 2013: 0.709
    - 2014: 0.703
    - 2011-2015: 0.706
  - Model calibration assessed via examination of plots of observed versus predicted values (graphs presented, along with R^2 statistics)
    - 2012: 0.992
    - 2013: 0.992
    - 2014: 0.990
    - 2011-2015: 0.997
- Some panel members expressed concern that DNR (Do Not Resuscitate) status is not accounted for in the measure
  - Information provided by the developer (01/03/2019)

- DNR is not currently captured in GWTG-Resuscitation and would be extremely difficult to capture, since DNR status can change throughout a patient's hospitalization. However, the denominator does include only patients who receive chest compressions and/or defibrillation. Since patients who are DNR would not receive chest compressions or defibrillation, they should not be included in the measure population.
- The RSSR measure adjusts for patients' case-mix severity, so that even if hospitals do have varying rates of DNR, the measure provides comparisons of how hospitals perform on the outcome of survival to discharge for their patients with an IHCA.

Standing Committee Action Item(s):

- The Standing Committee initially reviewed this measure in fall 2017, before it was withdrawn from the cycle. The Standing Committee should determine the reliability and validity of the measure score based on the testing submitted for this cycle.

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**

Combined Methods Panel Scientific Acceptability Evaluation

Measure Number: 3309

Measure Title: Risk-Standardized Survival Rate (RSSR) for In-Hospital Cardiac Arrest

Type of measure:

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

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

Data Source:

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

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

☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

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

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

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

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

   PANEL MEMBER 1: None

   PANEL MEMBER 4: None.

   PANEL MEMBER 5: No concerns.

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   PANEL MEMBER 1: Reliability was assessed through signal-to-noise analysis. Specifically, signal-to-noise ratio testing was conducted by fitting a hierarchical mixed effects model to derive the two shape parameters – alpha and beta; the model was built on a specified beta-binomial distribution. The two estimated model parameters were then used to calculate between-site (hospital-to-hospital) and within-site (hospital-specific) variances.
   PANEL MEMBER 2: Used a Signal-to -noise analysis. It is unclear if their interpretation of the result is correct.
   PANEL MEMBER 3: The developer uses the Beta-binomial model (Adams 2009) to estimate signal-to-noise
   PANEL MEMBER 4: An appropriate beta-binomial model was used.
   PANEL MEMBER 5: Signal-to -noise

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

   Submission document: Testing attachment, section 2a2.3

   PANEL MEMBER 1: Based on the reliability results, the risk-standardized survival rate (RSSR) measure has moderate reliability (overall mean reliability score of 0.76).

   PANEL MEMBER 2: For 2011-2015 → 0.76/0.78

   For 2013 → 0.70/0.72

   For 2014 → 0.67/0.68

PANEL MEMBER 3: The developer reports average or median reliability in excess of 0.6 or 0.7 at the facility level, which is moderate.

PANEL MEMBER 4: Based on the results of the signal to noise analysis, here was moderate reliability in the risk-standardized survival rate measure.

PANEL MEMBER 5: Moderate reliability.

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

   Submission document: Testing attachment, section 2a2.2

   ☒**Yes**

   ☐**No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒**Yes**

   ☐**No**

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

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

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

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

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

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

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

    PANEL MEMBER 1: The "moderate" reliability rating based on the signal-to-noise analysis results (see #7 above).

    PANEL MEMBER 2: The" reliability statistics were moderate, but not strong

    PANEL MEMBER 3: The developer reports average reliability. However, reliability is about noise (estimation error) relative to purpose. Therefore, reliability should be reported at the measured entity level in addition to the measure level (that is, the mean and the distribution of the reliability metric).

    PANEL MEMBER 4: Based on the results of the appropriate signal to noise analysis, here was moderate reliability in the risk-standardized survival rate measure.

    PANEL MEMBER 5: Identified moderate rates after testing.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    PANEL MEMBER 1: N/A

    PANEL MEMBER 2: No concerns

    PANEL MEMBER 3: None

PANEL MEMBER 4: None.

PANEL MEMBER 5: Not applicable.

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

Submission document: Testing attachment, section 2b4.

PANEL MEMBER 1: None

PANEL MEMBER 2: No concerns.

PANEL MEMBER 3: None

PANEL MEMBER 4: None.

PANEL MEMBER 5: Demonstrated wide variation in performance of the measure across facilities.

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

Submission document: Testing attachment, section 2b5.

PANEL MEMBER 1: N/A

PANEL MEMBER 2: Not applicable.

PANEL MEMBER 3: None

PANEL MEMBER 4: None.

PANEL MEMBER 5: Possibly, mild limitations of data set. For example, not being able to exclude DNR patients from model (as noted by panelist reviewer). Developers may consider for providing a rationale on expected magnitude of these limitations on the model.

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

Submission document: Testing attachment, section 2b6.

PANEL MEMBER 1: The potential effect of missing data could not be tested due to lack of data. However, given the very low rate of missingness in the relevant variables, I suspect that missing data could have significant impact on the measure score.

PANEL MEMBER 2: No concerns. The missing data were in small numbers.

PANEL MEMBER 3: None

PANEL MEMBER 4: None.

PANEL MEMBER 5: No concerns.

16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d. **Risk adjustment summary:**

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

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion?

☐ Yes ☒ No

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

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)

☒ Yes ☐ No

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

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

PANEL MEMBER 1: Multivariable hierarchical logistic regression model accounting for hospital-level clustering was used to predict in-hospital cardiac arrest. Once the final model as per the steps in 2b3.1.1 are determined, risk-standardized survival rates for each hospital are computed. The detailed steps are clearly delineated in 2b3.1.1.

PANEL MEMBER 2: Indicates that critical-care interventions are part of the risk-adjustment model. They are interventions before the cardiac arrest, but are not necessarily "not present" before the start of care.

PANEL MEMBER 3: Acceptable

PANEL MEMBER 4: As documented in the paper by Chan et al. (included in the original submission), the statistical risk adjustment model is appropriate. The results also indicate that the risk models are valid, predictive, descriptive, and are well-calibrated. I'm not sure that I agree with the decision not to adjust for race and other sociodemographic factors.

PANEL MEMBER 5: Left SDS factors in the model to reflect disparities in quality and, appropriately, aimed to identify a parsimonious model, which were both appropriate.

VALIDITY: TESTING

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

☒ Face validity

☐ Empirical validity testing of the measure score

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

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

Submission document: Testing attachment, section 2b2.2

PANEL MEMBER 1: Face validity was established through an expert panel comprising of experts from multiple specialties (=10) who would have potential stakes in this quality measure (see #2b1.3). This 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." The following rating scale from 1 to 5 was used, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; and 5= Strongly Agree.

PANEL MEMBER 2: Assessed face validity using a panel of 34 persons, all experts in the field.

PANEL MEMBER 3: No empirical validity testing of measure score

PANEL MEMBER 4: A formal expert process was used.

PANEL MEMBER 5: Consensus ratings from technical expert panel.

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

Submission document: Testing attachment, section 2b2.3

PANEL MEMBER 1: Approximately 71% said either "agree" or "strongly agree", implying that the measure can be considered valid.

PANEL MEMBER 2: 71% agreed or strongly agreed that the measure is a good measure of quality.

PANEL MEMBER 3: No empirical validity testing of measure score

PANEL MEMBER 4: Only 70% agreed with the statement that ""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." Also some of the 30% was explained, this is not a strong result.

PANEL MEMBER 5: Appropriate initial approach to assessing face validity; however, can open the measure up for input beyond expert panel (e.g., public comment).

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

    Submission document: Testing attachment, section 2b1.

    ☒**Yes**

    ☒**No**

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

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

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

    ☐**Yes**

    ☐**No**

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

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

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

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

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

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

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

    PANEL MEMBER 1: "Moderate" rating is because this is the maximum allowed in the absence of score-level testing.

    PANEL MEMBER 2: Assessed the face validity; found substantial agreement.

    PANEL MEMBER 3: No empirical validity testing of measure score or data elements. If this were a maintenance measure the rating would be insufficient.

    PANEL MEMBER 4: The only validity assessment was and those results were not strong

    PANEL MEMBER 5: The measure only received face validity testing, which is insufficient. , given it is an outcome measure seeking continued endorsement. Empirical validity testing is expected at this point in the measure's life cycle.

ADDITIONAL RECOMMENDATIONS

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

PANEL MEMBER 1: None.

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

2a1. Specifications:

- code sheet seems clear
- No concerns
- Reliability of the measure should be excellent.
- Measure specifications are defined and consistent with the evidence, but the specifications, particularly the numerator, need greater detail. I think the numerator should include "aged 18 years and older with in-hospital cardiac arrest who received chest compression and/or defibrillation," thus patients "aged 18 years and older with in-hospital cardiac arrest who received chest compression and/or defibrillation" who were alive at discharge.
- reliability testing was completed at the measure score level measured as the ratio of signal to noise. The mean/median scores were 0.76 and 0.78 at the hospital level for 2011-2015, however scores were slightly lower based on one year data.
- Data elements are clearly defined. No concerns for hospitals participating in the registry

2a2. Reliability testing:

- no concerns
- No
- I have no concern about reliability of the outcome measure but proper classification for diagnosis coding could be variable.
- Reliability testing was done using a signal-to-noise (SNR) analysis, specifically the Adams' beta-binomial method. Mean results in the 0.67-0.76 range, so moderate reliability.
- slighly less reliable over a shorter time frame.
- No concerns

2b1. Validity testing:

- no concerns
- Only face validity seems to be tested
- I have no concerns about the validity of testing the measure.
- Face validity testing done through TEP with agreement results at 71%, indicating moderate correlation. Overall moderate validity.
- Face validity showed 50% agree, 20.6% strongly agree, and 17.6% neither agreed or disagreed.
- Face validity only, but I don't have a significant concern about this.

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

- it is not clear how much data was missing, accumulating over the 9 variables in the model
- These do not seem to be used for validity testing
- Generally, I have no concerns about validity threats for this measure but in some cases a relatively small n may not yield statistically significant differences between institutions where a quality gap may exist.
- no serious concerns
- There are no significant threats to validity

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

- no other concerns
- none

- The exclusions seem logical and consistent with current evidence.
- No exclusions identified and no threats. Extensive risk adjustment done (but no social risk factors) with statistical risk model, and results demonstrate good model discrimination (>0.99).
- Risk adjustment variables are reasonable. They specifically did not include race in the risk adjustment. Does not include socioeconomic data.
- Exclusions are not a problem; risk adjustment appears to be valid

Criterion 3. Feasibility

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

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

- Data elements are generated or collected by and used by healthcare personnel during the provision of care and abstracted from a record by someone other than person obtaining original information.
    - Data collected through Get with the Guideline – Resuscitation Registry
- All data elements are in defined fields in electronic clinical data

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**

RATIONALE:

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility:

- no concerns
- No concerns
- Feasibility should only be limited by differential coding where the patients suitable for the measure outcome might not have been properly classified.
- Data collection obtained through registry data, specifically the Get with the Guideline – Resuscitation Registry. All data elements are in defined fields in electronic clinical data. Moderate feasibility.
- data elements are in defined electronic data
- The measure is feasible for GWTG participating hospitals

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.
    - 373 hospitals that are geographically diverse participate in the registry, and in 2017, 128 hospitals received public recognition.

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 results were vetted with the American Heart Association Research Committee chairs.
    - The feedback provided was incorporated in the final measure.
- Participants in the Get With The Guidelines-Resuscitation program have access to their data through the registry.
    - They also receive a separate feedback report of their risk-standardized in-hospital cardiac arrest results.

Additional Feedback: N/A

***Questions for the Committee**:*

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

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

RATIONALE:

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

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

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

Improvement results

- 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.
- 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. 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 potential harms.

Additional Feedback: N/A

***Questions for the Committee**:*

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

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

RATIONALE:

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

4a. Use:

- no concerns
- No concerns
- There are potential accountability and transparency issues related to public reporting of this measure.
- Measure is not currently reported publicly. It is used in the American Heart Association Get With The Guidelines-Resuscitation Professional Certification or Recognition Program, and it is currently in the early planning stages of a voluntary public reporting program for the Get With The Guidelines-Resuscitation program. No unexpected harms or potential harms identified. Moderate use and usability.
- used in recognition programs for qualilty improvement.
- GWTG has a public recognition program

4b. Usability:

- no concerns
- No concerns.
- This measure is useable and does contribute to high quality healthcare. The unintended consequences might be that some patients are misclassified intentionally to result in better scores for purposes of payment.
- Measure is not currently reported publicly. It is used in the American Heart Association Get With The Guidelines-Resuscitation Professional Certification or Recognition Program, and it is currently in the early planning stages of a voluntary public reporting program for the Get With The Guidelines-Resuscitation program. No unexpected harms or potential harms identified. Moderate use and usability.
- Used for quality improvement currently
- Hospitals are provided with feedback and benchmarking. No harms are identified

Criterion 5: Related and Competing Measures

Related or competing measures
The developer lists no related or competing measures.
Harmonization
N/A

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

5. Related and Competing:

- There are other related measures for cardiovascular mortality. All cause mortality for CVD patients can be harmonized but should be a distinct measure.
- No related measures identified.
- There are no related or competing measures

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 1/25/2019

- No comments or support/non-support choices have been submitted as of this date.

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 an 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 survival 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.

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_v4_08NOV18.docx

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

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

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:

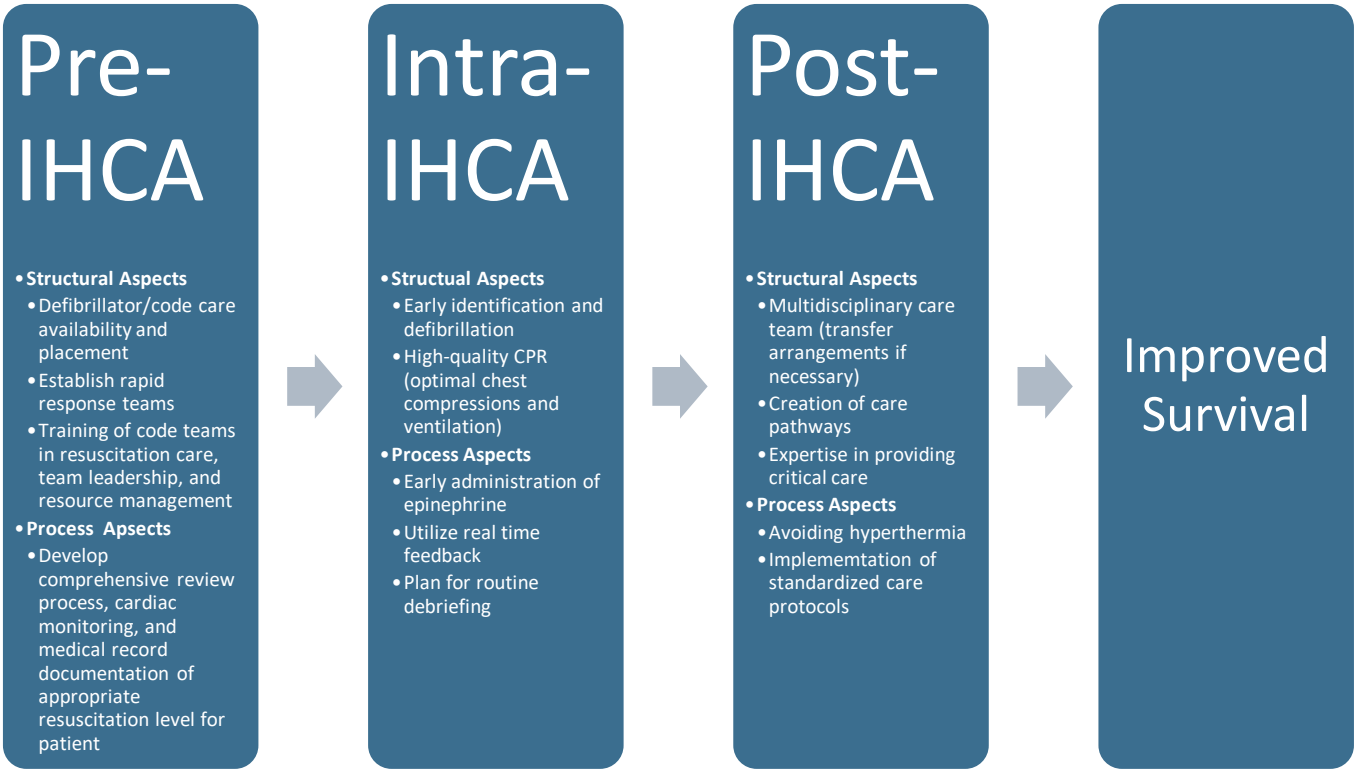
☐ Appropriate use measure:

☐ 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 outlines both structural factors and processes for all 3 phases of an in-hospital cardiac arrest (IHCA) hospitalization, all of which can influence the likelihood of 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). Recent studies have shown that increased duration of resuscitation attempt, prompt administration of epinephrine, and timely delivery of defibrillation can improve survival rates after an in-hospital cardiac arrest. 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 conducted longer resuscitations (median 25 minutes) had higher rates of return of spontaneous circulation and in-hospital survival, compared to hospitals that conducted shorter resuscitations (median 16 minutes) (Goldberger, et al., Lancet 2012). Another study found that hospitals with lower rates of delayed epinephrine treatment had higher survival rates for their patients with in-hospital cardiac arrest (Khera R, Chan PS, Donnino M, Girota S, Circulation, 2016). In a third study, while it is widely known that prompt delivery of defibrillation contributes to improved survival, the extent of benefit was unclear until a landmark quantified that patients with delays in defibrillation treatment (>2 minutes) have half the odds of survival to hospital discharge compared with promptly treated patients (Chan PS et al, NEJM, 2008). Additionally, survival rates after in-hospital cardiac arrest have improved with facility participation in the Get With The Guidelines-Resuscitation registry (from 16% up to 24% from 2010 to 2013) which could be associated with improved resuscitation care (Chan, National Academies, 2015).

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.

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)*

If a COMPOSITE (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 an 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 survival 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.*

As shown in section 2b4.2 of the testing attachment, performance scores are as follows:

2011- May 2015

Based on the sample of 326 hospitals during this time period, the mean performance risk-standardized survival rate was 24% (standard deviation of 5%), and the median performance rate was 24% (minimum rate of 11% and a maximum rate of 38%, with range of 27%).

2013

Based on the sample of 273 hospitals during this year, the mean performance risk-standardized survival rate was 25% (standard deviation of 5%), and the median performance rate was 25% (minimum rate of 9% and a maximum rate of 39%, with range of 30%).

2014

Based on the sample of 259 hospitals, the mean performance risk-standardized survival rate was 24% (standard deviation of 5%), and the median performance rate was 24% (minimum rate of 14% and a maximum rate of 40%, with range of 26%).

The range of performance demonstrated above suggests there is clinically meaningful variation across hospitals' risk-standardized survival rate for IHCA.

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 20 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 included in section 1.8 of 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. <u>For maintenance of endorsement:</u> Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

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

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

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

Patients who 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)*

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

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.)*

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

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. Hospitals with 20 or more cases of in-hospital cardiac arrest during the measurement period are identified as eligible for the measure.
2. Patients for inclusion are identified using inclusion criteria as described above (S.6 through S.9)
3. Patients meeting the numerator (S.4-S.5) are determined.
4. Variables for inclusion in risk adjustment are pulled.
5. 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-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. <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.)*

Not Applicable

2. Validity – See attached Measure Testing Submission Form

Resubmission_Document_3309_nqf_testing_attachment_08_16_2018_FINAL.docx

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

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

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.

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.

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

Measure Number (*if previously endorsed*)**:** 3309
Measure Title: Risk-Standardized Survival Rate for In-Hospital Cardiac Arrest
Date of Submission: 8/1/2018

Type of Measure:

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

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). NRCPR was started in 1999 to collect resuscitation data from

hospitals nationwide and create evidence-based guidelines for inpatient cardiac arrests. In 2010, the program was incorporated into Get With The Guidelines and enhanced to provide additional resources, tools and benefits, and now includes the following:

- performance of comparison with hospitals
- reduction of noncompliance and medical errors through data-driven peer review
- web-based data collection to fulfill Joint Commission standards and other requirements
- real-time assessment of resuscitation performance measures
- identification of quality improvement opportunities
- access to the most up-to-date research and scientific publications
- professional education opportunities, such as workshops and webinars
- clinical tools and resources
- quality improvement staff support in AHA's field offices
- a competitive advantage in the healthcare marketplace
- national and local recognition for hospital team program achievement

The Get With The Guidelines-Resuscitation program is administered by the American Heart Association/American Stroke Association.

1.3. What are the dates of the data used in testing? Data extracted from the Get With the Guidelines Resuscitation registry were used to describe the patient case mix and eligible patient population (initial derivation cohort). Initial model derivation and validation results used data from 01/01/2007 to 12/31/2010. In the current application to NQF, we have performed prospective validation and reliability testing using data from 01/2011 to 05/2015 (prospective validation cohort). Additionally, we conducted reliability analyses for years 2013 and 2014 separately.31/2010. In the current application to NQF, we have performed prospective validation and reliability testing using data from 01/2011 to 05/2015 (prospective validation cohort). Additionally, we conducted reliability analyses for years 2013 and 2014 separately.Data extracted from the Get With the Guidelines Resuscitation registry were used to describe the patient case mix and eligible patient population (initial derivation cohort). Initial model derivation and validation results used data from 01/01/2007 to 12/

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*)

For the 2011 - May 2015 analyses:

For the main analyses for NQF submission, a total of 326 hospitals reported on survival outcomes, collected clinical data for risk adjustment, and had cases of in-hospital cardiac arrest between 2011 and May 2015. The

average number of quality reporting events per hospital was 190, for a total of 61,934 cardiac arrest events, and 14,782 (23.9%) patients survived to hospital discharge. Across the 326 hospitals, the range of cardiac arrest quality reporting events was 1 to 1222, and the range for number of patients surviving to hospital discharge was 0 to 344.

We also conducted reliability analyses for 1-year time frames for years 2013 and 2014:

For the 2013 analyses**:**

A total of 273 hospitals reported on survival outcomes, collected clinical data for risk adjustment, and had cases of in-hospital cardiac arrest in 2013. The average number of quality reporting events per hospital was 66, for a total of 17,992 cardiac arrest events, and 4417 (24.5%) patients survived to hospital discharge. Across the 273 hospitals, the range of cardiac arrest quality reporting events was 1 to 360, and the range for number of patients surviving to hospital discharge was 0 to 121. For the 1-year reliability analyses for year 2013, we restricted the analyses to the 206 hospitals which had a minimum number of 20 quality reporting events.

For the 2014 analyses**:**

A total of 259 hospitals reported on survival outcomes, collected clinical data for risk adjustment, and had cases of in-hospital cardiac arrest in 2014. The average number of quality reporting events per hospital was 67, for a total of 17,244 cardiac arrest events, and 4163 (24.1%) patients survived to hospital discharge. Across the 259 hospitals, the range of cardiac arrest quality reporting events was 1 to 409, and the range for number of patients surviving to hospital discharge was 0 to 124. For the 1-year reliability analyses for year 2014, we restricted the analyses to the 200 hospitals which had a minimum number of 20 quality reporting events.

Reliability analyses in this submission were conducted using data from the entire prospective validation cohort (2011 to May 2015) and from the 12-month periods of 2013 and 2014. In the table below, we provide summary characteristics of the hospitals in all 3 time periods.

Hospital Characteristics*

Teaching Status	**2011 - May 2015 (n = 326)**	**2013 (n = 273)**	**2014 (n= 259)**
Major teaching	93 (29.3%)	84 (31.3%)	73 (28.7%)
Minor teaching	94 (29.7%)	76 (28.4%)	76 (29.9%)
Non-teaching	130 (41.0%)	108 (40.3%)	105 (41.3%)
Missing (.)	9	5	5
Total Beds			
<100	21 (6.7%)	18 (6.8%)	14 (5.6%)
100-199	56 (17.9%)	46 (17.4%)	45 (17.9%)
200-249	26 (8.3%)	22 (8.3%)	21 (8.4%)
250-299	38 (12.2%)	29 (11.0%)	28 (11.2%)
300-349	30 (9.6%)	27 (10.2%)	28 (11.2%)
350-499	67 (21.5%)	55 (20.8%)	52 (20.7%)
500+	74 (23.7%)	67 (25.4%)	63 (25.1%)
Missing (.)	14	9	8
Level of Trauma Center			
Regional	81 (39.7%)	71 (41.0%)	64 (39.3%)
Community	76 (37.3%)	60 (34.7%)	58 (35.6%)
Rural	45 (22.1%)	40 (23.1%)	40 (24.5%)
Other	2 (1.0%)	2 (1.2%)	1 (0.6%)
Missing (.)	122	100	96
Census Division Region			

Teaching Status	2011 - May 2015 (n = 326)	2013 (n = 273)	2014 (n= 259)
North Mid Atlantic	69 (21.6%)	62 (23.0%)	53 (20.7%)
South Atlantic & Puerto Rico	70 (21.9%)	61 (22.6%)	58 (22.7%)
North Central	70 (21.9%)	54 (20.0%)	54 (21.1%)
South Central	49 (15.3%)	43 (15.9%)	43 (16.8%)
Mountain/Pacific	62 (19.4%)	50 (18.5%)	48 (18.8%)
Missing (.)	6	3	3
Location			
Rural	25 (7.9%)	21 (7.8%)	18 (7.1%)
Urban	292 (92.1%)	247 (92.2%)	236 (92.9%)
Missing (.)	9	5	5

* Summary table provides hospital characteristics for all hospitals during each time period, regardless of case volume. Percentages reflect hospitals without missing data for each characteristic.

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 initial derivation cohort for the model for risk-standardized survival included 32,560 patients with an in-hospital cardiac arrest between 2007 and 2010, and the initial validation cohort comprised 16,281 patients with an in-hospital cardiac arrest during this same time period. These results were published in the Journal of the American College of Cardiology [1]. For this application, we have conducted a *prospective* validation of the initial model for risk-standardized survival after in-hospital cardiac arrest using data on 61,934 patients with in-hospital cardiac arrest between January of 2011 and May of 2015. Additionally, we conducted reliability analyses for this entire prospective time period, as well as for years 2013 and 2014 separately, given that our proposed measure will be an annual (1-year) measure.

1. Chan PS, Berg RA, Spertus JA, et al. for the AHA GWTG-Resuscitation Investigators. Risk-standardizing survival for in-hospital cardiac arrest to facilitate hospital comparisons. J Am Coll Cardiol 2013;62:601–9. doi:10.1016/j.jacc.2013.05.051

The table below and on the next page describes the patient case mix (demographics and pre-existing conditions) for the initial derivation (2007-2010), initial validation (2007-2010) and prospective validation cohorts (2011-May 2015). The patient populations in each of these 3 cohorts were very similar in case-mix.

	Population Clinical Characteristics		
	Initial Derivation Cohort 2007-2010	Initial Validation Cohort 2007-2010	Prospective Validation Cohort 2011-May 2015
	(n = 32,560)	(n = 16,281)	(n=61,934)
Demographics			
Age, mean ± standard deviation	65.6 ± 16.1	65.6 ± 16.0	65.2 ± 15.9
Male sex	18,996 (58.3%)	9,500 (58.4%)	36,241 (58.5%)
Race			
White	22,576 (69.3%)	11,337 (69.6%)	42,580 (68.8%)
Black	6,678 (20.5%)	3,288 (20.2%)	14,138 (22.8%)
Other	1,268 (3.9%)	618 (3.8%)	1,530 (2.5%)
Unknown	2,038 (6.3%)	1,038 (6.4%)	3,686 (6.0%)
Hispanic ethnicity	2,254 (6.9%)	1,060 (6.5%)	2,780 (4.5%)

	Population Clinical Characteristics		
	Initial Derivation Cohort 2007-2010	**Initial Validation Cohort 2007-2010**	**Prospective Validation Cohort 2011-May 2015**
Pre-Existing Conditions			
Respiratory insufficiency	13,301 (40.9%)	6,640 (40.8%)	26,527 (42.8%)
Renal insufficiency	10,850 (33.3%)	5,358 (32.9%)	21,336 (34.4%)
Diabetes mellitus	10,001 (30.7%)	4,928 (30.3%)	19,652 (31.7%)
Hypotension	8,413 (25.8%)	4,308 (26.5%)	14,645 (23.6%)
Heart failure during admission	5,370 (16.5%)	2,678 (16.4%)	9,527 (15.4%)
Prior heart failure	6,278 (19.3%)	3,094 (19.0%)	12,971 (20.9%)
Myocardial infarction during admission	5,184 (15.9%)	2,501 (15.4%)	8,807 (14.2%)
Prior Myocardial infarction	4,791 (14.7%)	2,319 (14.2%)	8,389 (13.5%)
Metabolic or electrolyte abnormality	4,765 (14.6%)	2,280 (14.0%)	10,640 (17.2%)
Septicemia	5,519 (17.0%)	2,777 (17.1%)	10,550 (17.0%)
Pneumonia	4,342 (13.3%)	2,239 (13.8%)	8,445 (13.6%)
Metastatic or hematologic malignancy	4,046 (12.4%)	1,997 (12.3%)	7,108 (11.5%)
Hepatic insufficiency	2,474 (7.6%)	1,175 (7.2%)	4,434 (7.2%)
Baseline depression in CNS function	3,640 (11.2%)	1,853 (11.4%)	5,449 (8.8%)
Acute CNS non-stroke event	2,250 (6.9%)	1,139 (7.0%)	3,797 (6.1%)
Acute stroke	1,234 (3.8%)	605 (3.7%)	2,266 (3.7%)
Major trauma	1,399 (4.3%)	668 (4.1%)	2,853 (4.6%)
Characteristics of arrest			
Cardiac arrest rhythm			
Asystole	10,997 (33.8%)	5,491 (33.7%)	17,893 (28.9%)
Pulseless electrical activity	15,327 (47.1%)	7,653 (47.0%)	33,240 (53.7%)
Ventricular fibrillation	3,691 (11.3%)	1,862 (11.4%)	6,149 (9.9%)
Pulseless ventricular tachycardia	2,545 (7.8%)	1,275 (7.8%)	4,652 (7.5%)
Location of cardiac arrest			
Intensive care unit	15,780 (48.5%)	7,809 (48.0%)	30,084 (48.6%)
Monitored unit	5,034 (15.5%)	2,539 (15.6%)	9,442 (15.2%)
Non-monitored unit	5,632 (17.3%)	2,824 (17.3%)	9,477 (15.3%)
Emergency room	3,307 (10.2%)	1,687 (10.4%)	7,072 (11.4%)
Procedural or surgical area	2,132 (6.5%)	1,073 (6.6%)	4,662 (7.5%)
Other	675 (2.1%)	349 (2.1%)	1,197 (1.9%)
Interventions in Place			
Mechanical ventilation	10,747 (33.0%)	5,422 (33.3%)	20,604 (33.3%)
Intravenous vasopressor	9,549 (29.3%)	4,800 (29.5%)	14,177 (22.9%)
Dialysis	1,163 (3.6%)	598 (3.7%)	1,687 (2.7%)

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.

Although data from 2007-2010 in the Get With the Guidelines Resuscitation registry were used to derive the initial risk-standardization model (initial derivation cohort) and validate that model (initial validation cohort), in the current application to NQF, we have performed prospective validation and reliability testing using data from 01/2011 to 05/2015 (prospective validation cohort). Additionally, we conducted reliability analyses for 1-year periods for years 2013 and 2014 separately. There were no significant differences between the 2013 or 2014 study population as compared with the larger prospective validation cohort (2011 to May 2015), of which it is part., we have performed prospective validation and reliability testing using data from 01/2011 to 05/2015 (prospective validation cohort). Additionally, we conducted reliability analyses for 1-year periods for years 2013 and 2014 separately. There were no significant differences between the 2013 or 2014 study population as compared with the larger prospective validation cohort (2011 to May 2015), of which it is part.Although data from 2007-2010 in the Get With the Guidelines Resuscitation registry were used to derive the initial risk-standardization model (initial derivation cohort) and validate that model (initial validation cohort), in the current application to NQF

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 risk-standardized survival rate measure for in-hospital cardiac arrest is a hospital-level measure, race-specific survival was not assessed at the patient-level. Instead, we examined the influence of race on the risk-standardized survival rate measure by dividing study hospitals into quartiles of patients with in-hospital cardiac arrest who were of black race. Across hospitals, the median percentage of in-hospital cardiac arrest patients of black race was 11% (IQR: 4% to 27%). In the table below, we outline that hospitals with the lowest proportion of black patients with in-hospital cardiac arrest (quartile 1) had a higher rate of unadjusted and risk-standardized survival for cardiac arrest as compared with hospitals that had the highest proportion of black patients (quartile 4), suggesting some degree of disparity in risk-standardized survival rates by hospital racial composition despite adjustment for patient case-mix severity (see table below).

In our models, we deliberately chose to not include race/ethnicity as a covariate in deriving our model for risk-standardized survival for in-hospital cardiac arrest so as to not mask disparities in care and outcomes for this condition by race. Including race in the model would have, in effect, made it more acceptable for hospitals with higher proportions of black patients with in-hospital cardiac arrest to have lower survival rates compared with other hospitals.

	Hospital Quartile By Proportion of Black IHCA Patients					
	Least Black			**Most Black**	**All Hospitals n = 288**	**P**
	Q1; n=72	**Q2; n = 72**	**Q3; n = 72**	**Q4; n = 72**		
Observed Rate						< 0.001
Mean ± SD	0.26 ± 0.09	0.24 ± 0.08	0.24 ± 0.07	0.20 ± 0.07	0.24 ± 0.08	
Median (IQR)	0.27(0.20, 0.31)	0.23 (0.19, 0.28)	0.24 (0.20, 0.28)	0.20 (0.17, 0.23)	0.23 (0.19, 0.28)	
Risk-standardized survival rate						0.002
Mean ± SD	0.25 ± 0.05	0.24 ± 0.05	0.25 ± 0.06	0.22 ± 0.05	0.24 ± 0.05	
Median (IQR)	0.25 (0.22, 0.29)	0.24 (0.20, 0.28)	0.25 (0.21, 0.29)	0.23 (0.19, 0.26)	0.24 (0.21, 0.28)	

Abbreviation: IHCA, in-hospital cardiac arrest

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 and the noise is the total variability in measured 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 mixed effects model to derive the two shape parameters – alpha and beta; the model was built on a specified beta-binomial distribution. 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:

$$\text{Reliability} = \frac{\text{(hospital-to-hospital variance)}}{\text{((hospital-to-hospital variance) + (hospital-specific variance))}}$$

$$\text{Between-site (or hospital-to-hospital) variance} = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

$$\text{Within-site (or hospital-specific) variance} = \frac{\hat{p}_i(1-\hat{p}_i)}{n_{is}}$$

Where,

$\hat{p}_i$ = the proportion of patients who survived to discharge at hospital i.

n_{is} = the total number of cardiac arrest events at hospital i.

α, β = shape parameters

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*)

We calculated reliability statistics for the entire prospective validation period, and then for years 2013 and 2014 separately.

For the entire prospective validation period of 2011 to May 2015, the signal-to noise analysis resulted in a mean reliability score of 0.76 and a median reliability score was 0.78 for hospitals eligible for the measure. For the 1-year period of 2013, the mean and median reliability scores were 0.70 and 0.72, respectively. And for year 2014, the mean and median reliability scores were 0.67 and 0.68, respectively.

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?*)

There was moderate reliability in the risk-standardized survival rate measure, based on the results of the signal to noise analysis.

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." The following rating scale from 1 to 5 was used, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; and 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 (Strongly Disagree) - 1 response (2.9%)

2 (Disagree) - 3 responses (8.8%)

3 (Neither Agree nor Disagree) - 6 responses (17.6%)

4 (Agree) - 17 responses (50%)

5 (Strongly Agree) - 7 responses (20.6%)

The summary of the expert panel ratings of the validity statement were as follows:

N = 34; Mean rating = 3.76 with 71% of respondents with either an 'agree' or a 'strongly agree' response that this measure can accurately distinguish good and poor quality. Respondents had an opportunity to provide additional comments as part of the face validity survey.

(a) Of 24 persons with either an 'agree' or 'strongly agree' response, 12 provided comments. In the strongly agree category, the majority of commenters felt that the risk adjustment strategy was well vetted, scientifically sound, and addressed appropriately performed model validation. Those in the agreed category noted that the measure does not address some potential confounders (e.g. racial/ethnic variation, regional differences in outcomes, socioeconomic status).

(b) Of 6 persons with either a 'neither agree or disagree' response, 5 provided comments. Some noted that risk adjustment was not their expertise and another noted they were a non-clinician. One clinician noted that the measure does not account for DNR rates across hospitals and may have some unintended effect.

(c) Of 4 persons with either a 'disagree' or 'strongly disagree' response, 3 provided comments. One commenter that strongly disagreed noted that the measure was reasonable and pragmatic but the denominator may have limitations (e.g. cannot risk adjust for DNR or selection of healthier patients for

resuscitation). Those that disagreed noted that, despite risk adjustment, other patient factors and race may remain confounders.

In the table below, the specialty of the expert panel is summarized. Our face validity survey was administered to a diverse group of experts with the goal of reducing bias.

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 9 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.

There are several steps taken to derive and calculate a hospital's risk standardized survival rate for in-hospital cardiac arrest. These include:

1) Create a model for predictors of in-hospital cardiac arrest. Since patients at a given hospital with in-hospital cardiac arrest will have correlated outcomes, we used multivariable hierarchical logistic regression, wherein patients are nested within hospitals in the model and hospitals are modeled as random effects to account for clustering 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) were considered for model inclusion. We considered almost all variables (except race) as potential predictors in the model.
3) An initial "full" model is generated for significant predictors of the outcome of survival to discharge.
4) Within this initial "full" model, we then work to create a parsimonious model, by sequentially eliminating predictors with the smallest contribution to the model. This was an iterative process of reducing the number of covariates in the model until a "reduced" model with no less than 95% of the initial "full" model's predictive ability is achieved. In essence, this creates a model with many fewer variables while maintaining nearly the same predictive (discriminative) ability as the "full" model. The validated parsimonious model is comprised of 9 key variables.
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" model is confirmed, risk-standardized survival rates for each hospital are computed. 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 risk-standardized survival rate higher than the weighted mean, and one with a ratio < 1 would have a risk-standardized survival rate below the weighted mean.
7) The expected survival number (denominator) is 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.

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, emergency room)
3) Initial cardiac rhythm
4) Pre-existing conditions/present on arrival conditions (i.e. heart failure, septicemia)
5) Critical-care interventions in place prior to the arrest (i.e. mechanical ventilation, intravenous vasopressor support)

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.

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 no less than 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 risk-standardized survival rate measure does not adjust for race in the model. From prior work, we know that black patients have lower survival for in-hospital cardiac arrest than white patients (Chan et. al., 2009). Adjusting for race in the model would, in effect, would make "acceptable" the reality that black patients with in-hospital cardiac arrest have lower survival than white patients. It is our belief that, while race is a significant predictor of survival outcomes for IHCA, the risk-standardized survival rate measure should not provide a rationale for accepting existing disparities in care.

In this NQF submission, we do conduct additional analyses illustrating that hospitals with a greater proportion of black IHCA patients have markedly lower survival for all (black and white) their patients with IHCA (see 1.8). This subgroup analysis, stratified by the proportion of black patients with in-hospital cardiac arrest at each hospital, underscores the vast disparities in survival for this condition and highlights important care gaps in the contemporary management of this condition.

In summary, then, the model for our risk-standardized survival rate measure does not adjust for race, in order to not make acceptable existing disparities in in-hospital cardiac arrest survival by race. The risk-standardized survival rates, however, can be aggregated based on the proportion of patients with IHCA at a given hospital that are of black race, to examine whether racial disparities in survival for IHCA do exist.

Finally, GWTG-Resuscitation does not collect other socioeconomic variables of risk, such as income, employment status, or educational level. Were these patient-level factors available, we would not have included them in the derivation of the risk-standardized survival rate model for the same reason as we did not include race in the model, as it would provide an exception to worse care for patients of lower socioeconomic status. Nonetheless, we do not have any reason to believe that patient-level socioeconomic status would directly impact survival from IHCA, as clinicians responding to an emergency such as a cardiac arrest would not be aware of a patient's social or economic risk factors and their treatment decisions during an acute resuscitation would not influenced by these considerations. It is possible that IHCA patients of lower socioeconomic status may be more frequently treated at hospitals with worse survival outcomes. To the extent that this is the case, the model for our risk-standardized survival rate measure should not adjust for socioeconomic factors in order to not make acceptable existing disparities in in-hospital cardiac arrest survival by socioeconomic factors.

Chan PS, Nichol G, Krumholz HM, Spertus JA, Jones PG, Peterson ED, Rathore SS, Nallamothu BK, for the American Heart Association National Registry of Cardiopulmonary Resuscitation (NRCPR) Investigators. Racial differences after in-hospital cardiac arrest. JAMA. 2009;302(11):1195-1201.

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?

The following table lists the risk factors that were included in the final parsimonious model, along with their estimates, ORs, CI and P-Values. For our model reduction methodology, please see paragraph 3 of section 2b3.3a above.

Final Reduced Model of Significant Predictors for Survival to Discharge

Prospective Validation Cohort (2011- May 2015)

Predictor	Beta Estimate	Odds Ratio	Lower CI	Upper CI	P Value
Age (years)					
<50		Reference			Reference
50 to <60	0.046	1.05	0.98	1.12	0.18
60 to <70	-0.115	0.89	0.84	0.95	0.00
70 to <80	-0.342	0.71	0.67	0.76	<.0001
≥ 80	-0.680	0.51	0.47	0.54	<.0001
Location					
Non-monitored		Reference			Reference
Intensive care	0.188	1.21	1.13	1.29	<.0001
Monitored unit	0.318	1.37	1.28	1.48	<.0001
Emergency room	0.201	1.22	1.13	1.32	<.0001
Procedural area	0.941	2.56	2.36	2.79	<.0001
Other	0.439	1.55	1.35	1.78	<.0001
Arrest Rhythm					
Asystole		Reference			Reference
Pulseless electrical activity	0.005	1.00	0.96	1.05	0.85
Ventricular fibrillation	1.105	3.02	2.83	3.23	<.0001
Pulseless VT	1.072	2.92	2.71	3.14	<.0001
Hepatic Insufficiency	-0.587	0.56	0.51	0.61	<.0001
Hypotension	-0.513	0.60	0.57	0.63	<.0001
Septicemia	-0.350	0.70	0.66	0.75	<.0001
Metastatic Malignancy	-0.703	0.49	0.46	0.53	<.0001
Mechanical Ventilation	-0.483	0.62	0.59	0.65	<.0001
Continuous Vasopressor	-0.702	0.50	0.47	0.53	<.0001

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 by examining observed vs. predicted plots.

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, 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 prospective validation cohort (2011 to May 2015), model discrimination was similar (C-statistic of 0.707).

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

Below, we describe our risk model calibration statistics for the prospective validation cohort (2011 to May 2015). We describe them for year 2012, 2013, and 2014 separately, and 2011-May 2015 as a whole.

2012 DATA

Using data from this year alone, we re-developed and confirmed that the parsimonious model would be comprised of the same 9 predictors (c-statistic 0.694). The model for 2012 data calibrated well, with an R^2 of 0.992 (below). The discrimination and validation analyses using 2012 data prospectively validates the initial risk-standardized survival rate model using data between 2007 and 2010.

2013 DATA

Using data from this year alone, we re-developed and confirmed that the parsimonious model would be comprised of the same 9 predictors (c-statistic 0.709). The model for 2013 data also calibrated well, with an R^2 of 0.992 (below). The discrimination and validation analyses using 2013 data prospectively validates the prior risk-standardized survival rate model using data between 2007 and 2010.

2014 DATA

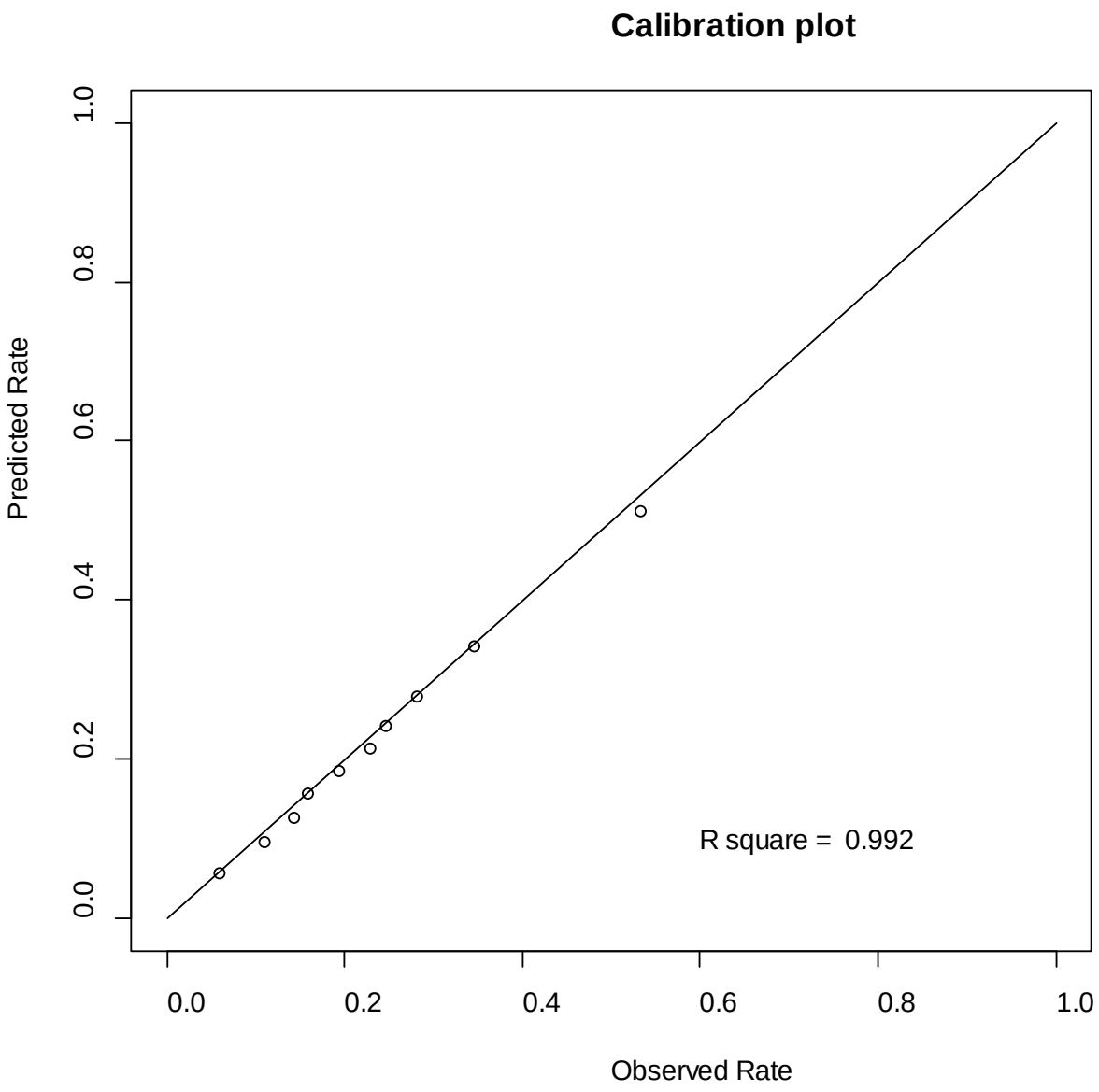
Using data from this year alone, we re-developed and confirmed that the parsimonious model would be comprised of the same 9 predictors (c-statistic 0.703). The model for 2014 data also calibrated well, with an R^2 of 0.99 (below). The discrimination and validation analyses using 2014 data prospectively validates the prior risk-standardized survival rate model using data between 2007 and 2010.

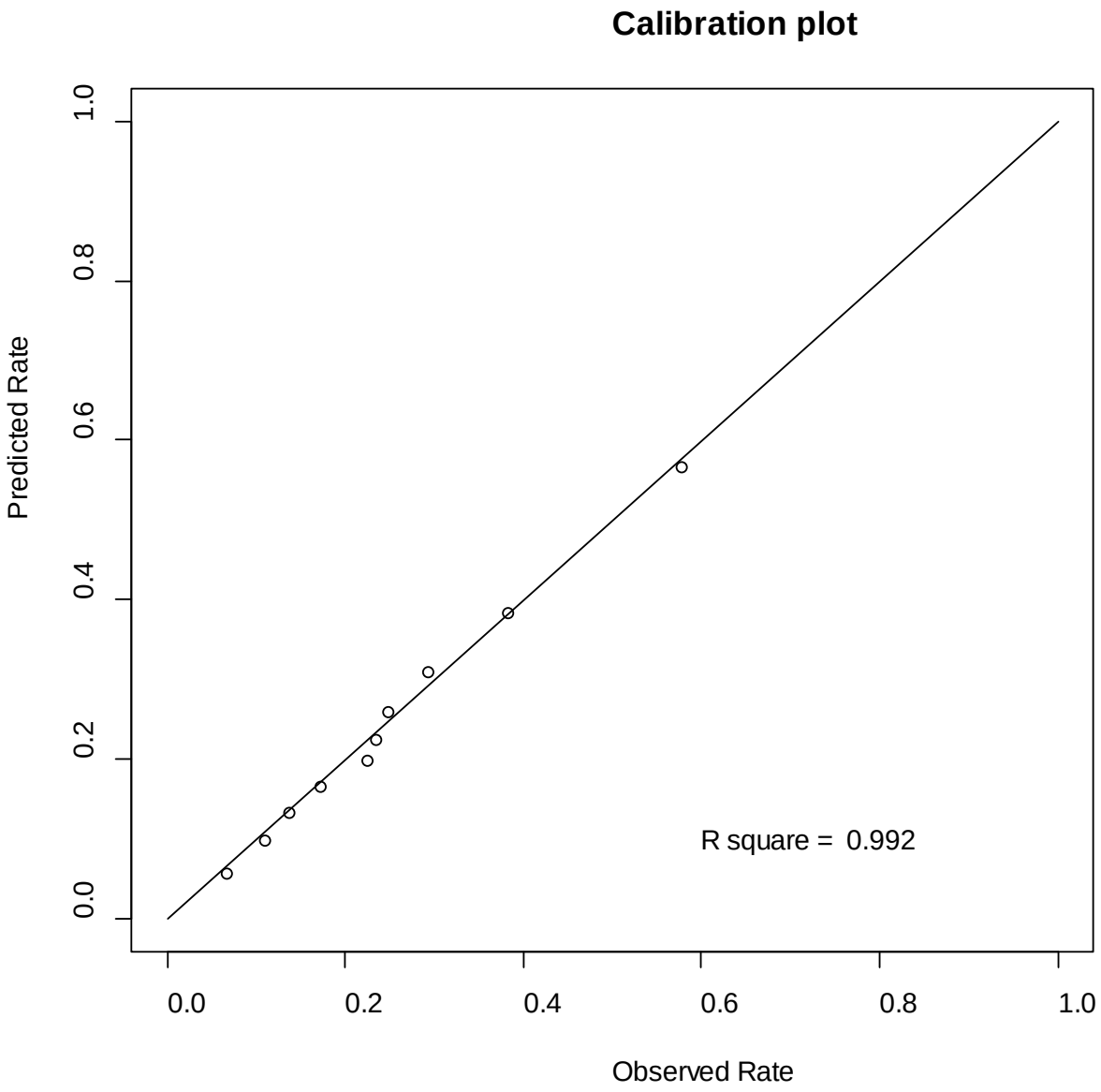
2011- May 2015

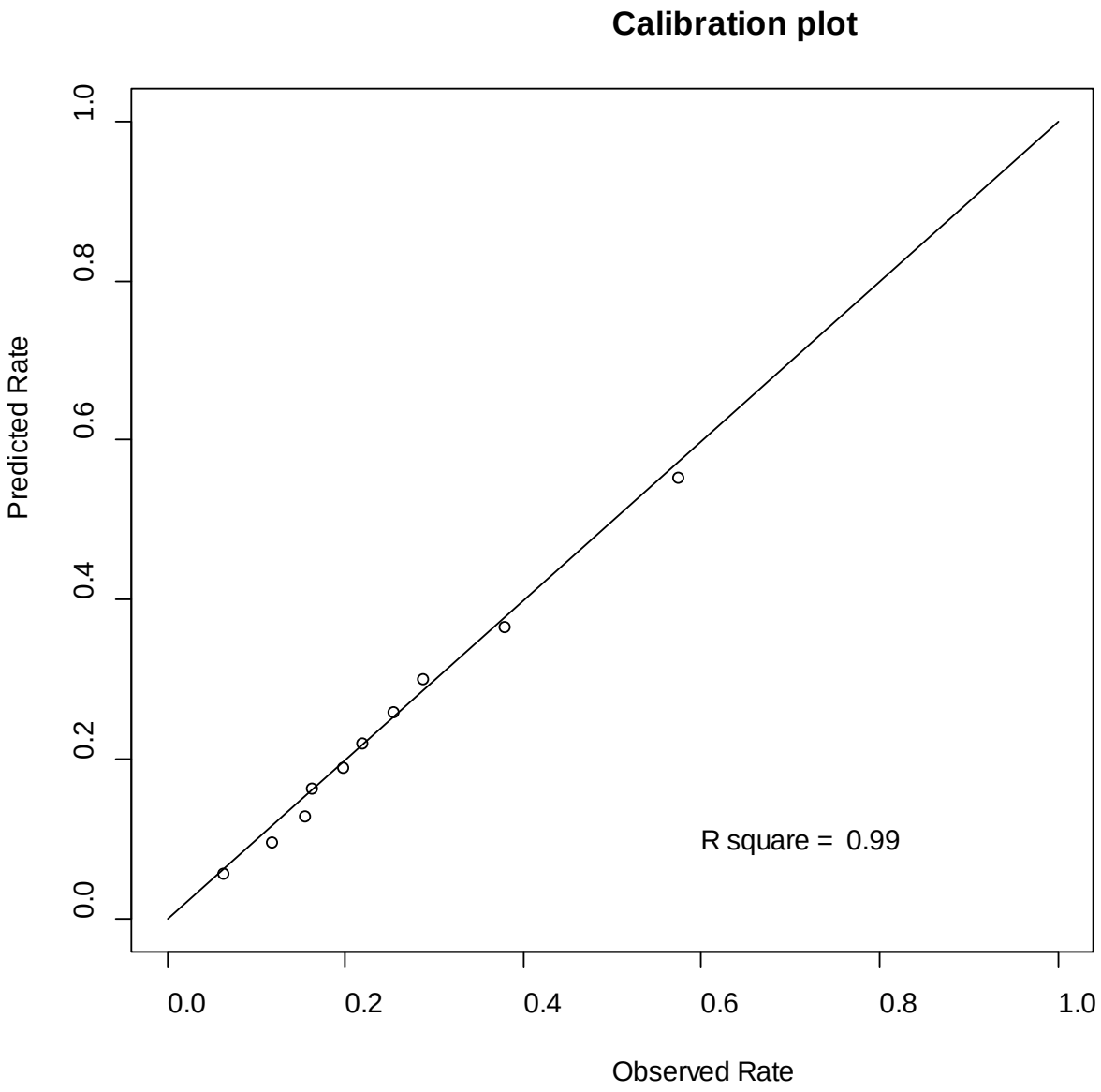
Using data from this entire time period, we re-developed and confirmed that the parsimonious model would be comprised of the same 9 predictors (c-statistic 0.706). The model using 2011-2015 data also calibrated well, with an R^2 of 0.997 (below). The discrimination and validation analyses using combined 2011-2015 data prospectively validates the initial risk-standardized survival rate model, which used data between 2007 and 2010.

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

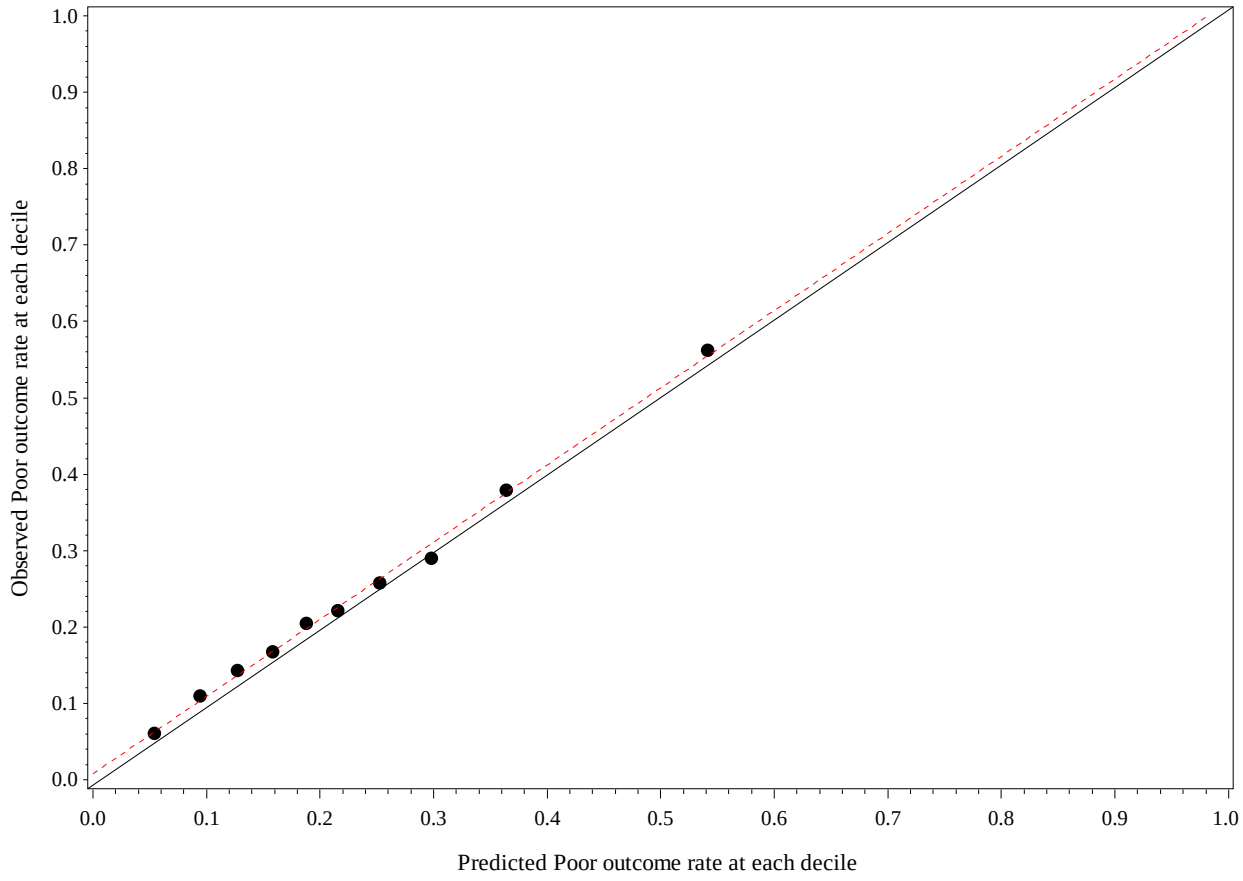
2012 DATA




Calibration plot
Predicted Rate
Observed Rate
0.0
0.2
0.4
0.6
0.8
1.0
R square = 0.992


Calibration plot
Predicted Rate
0.0
0.2
0.4
0.6
0.8
1.0
R square = 0.99
Observed Rate
0.0
0.2
0.4
0.6
0.8
1.0

2011- May 2015 DATA



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*)

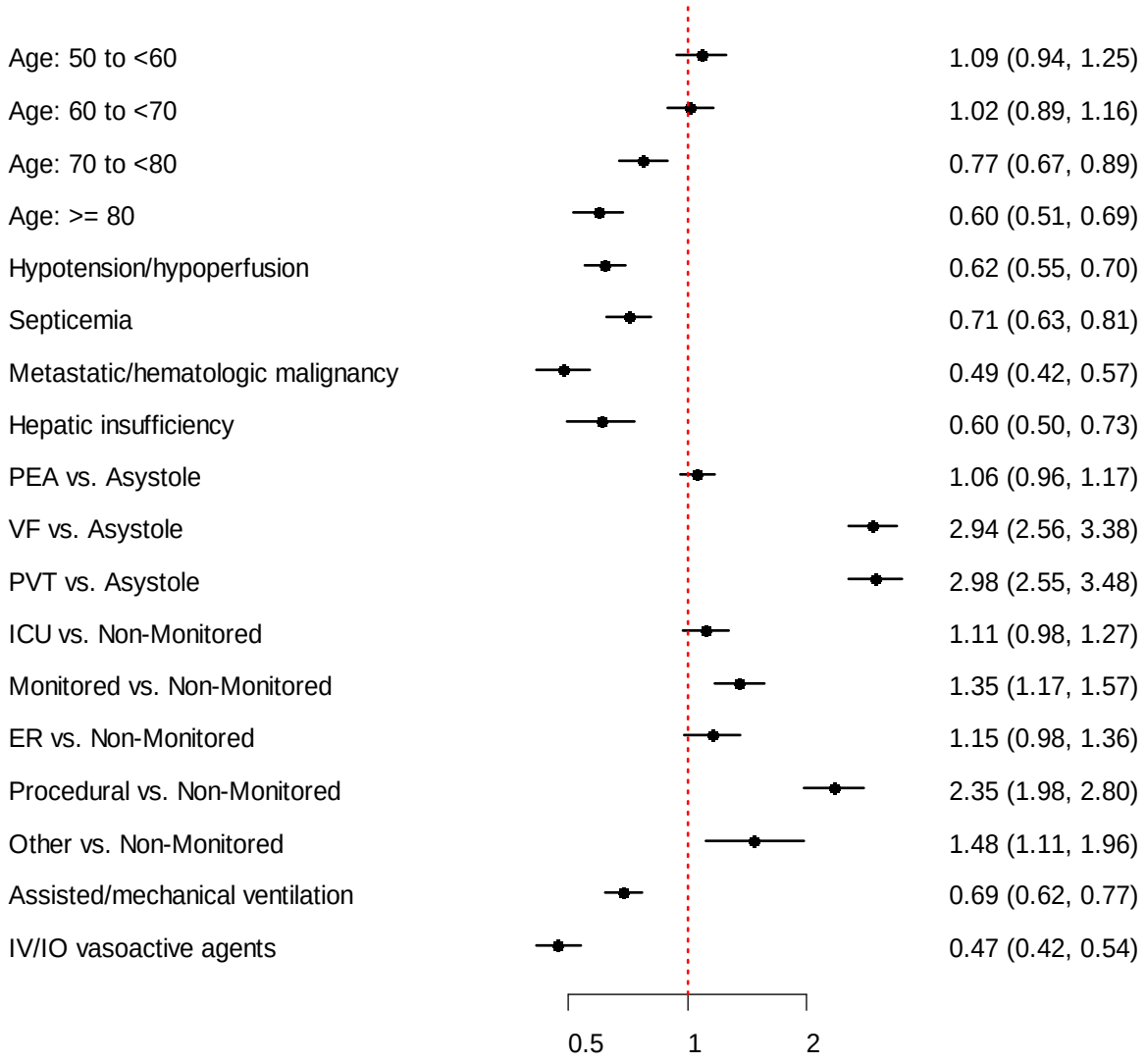
The results above indicate that the risk models are valid, predictive, descriptive, and are well-calibrated.

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*)

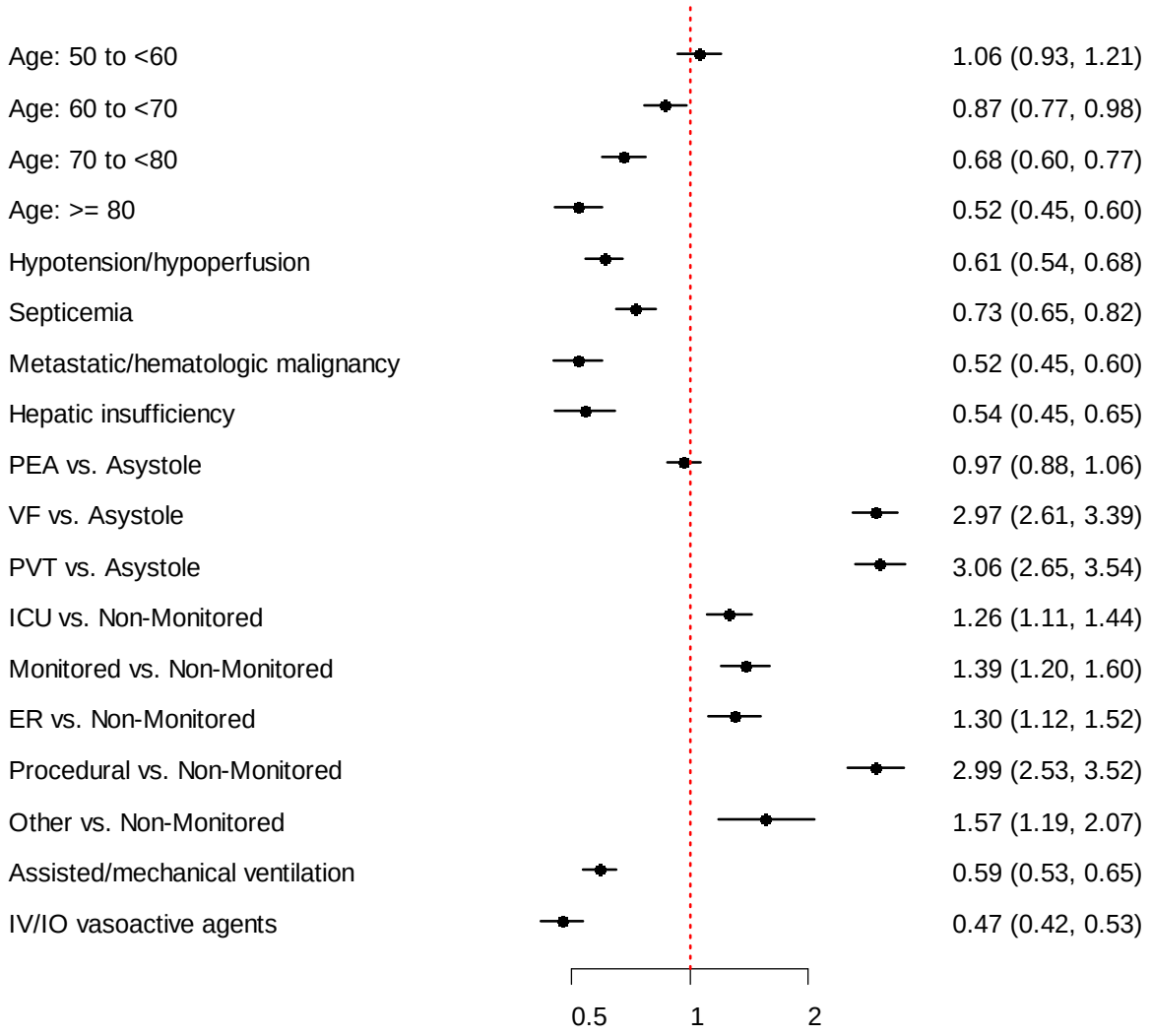
Forest plots

We provide below forest plots for the parsimonious model for the following years: 2012, 2013, and 2014. These plots illustrate the statistical significance of each variable, compared to its reference.

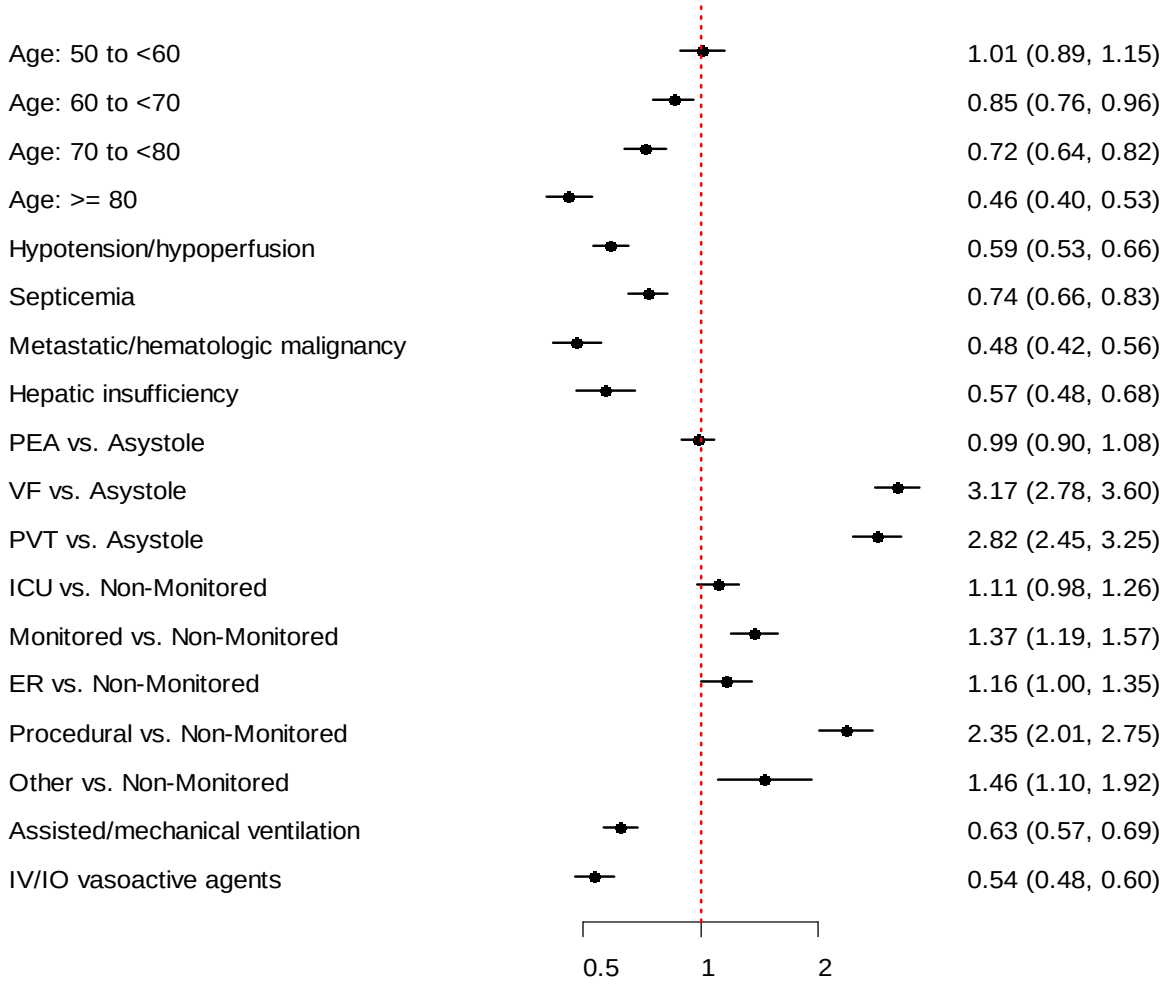
2012 Forest Plot



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.

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*)

2011- May 2015

Based on the sample of 326 hospitals during this time period, the mean performance risk-standardized survival rate was 24% (standard deviation of 5%), and the median performance rate was 24% (minimum rate of 11% and a maximum rate of 38%, with range of 27%).

2013

Based on the sample of 273 hospitals during this year, the mean performance risk-standardized survival rate was 25% (standard deviation of 5%), and the median performance rate was 25% (minimum rate of 9% and a maximum rate of 39%, with range of 30%).

2014

Based on the sample of 259 hospitals, the mean performance risk-standardized survival rate was 24% (standard deviation of 5%), and the median performance rate was 24% (minimum rate of 14% and a maximum rate of 40%, with range of 26%).

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 demonstrated above suggests there is clinically meaningful variation across hospitals' risk-standardized survival rate for IHCA.

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 was missing on survival to discharge for <1% of all patients in the registry and results were not imputed given that survival to discharge is the outcome variable of interest for the proposed measure. Otherwise, data were not missing for other covariates except for race (~5% to 6%) (see below), and as explained earlier in this

submission, we did not adjust for race in the models to avoid providing for exceptions for worse care in hospitals with a larger proportion of patients of black race.

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*)

1) Data on survival to discharge is missing in <0.1% of patients in the registry. Those patients with missing data on the outcome were excluded from the initial development and subsequent validation of the risk-standardized survival rate measure.

2) Data on patient demographics (age, sex) has no missing data. If the age/sex fields are not included in the submission, the electronic registry file cannot be submitted and the online data platform prompts the person entering the data to complete those fields. Therefore, the file cannot be submitted without completion of the age/sex fields. Data on race is missing in about 5-6% of patients. However, for reasons presented earlier in this submission, race is not included as a variable for risk adjustment.

3) Data on other patient variables (pre-existing comorbidities and conditions, and interventions at the time of cardiac arrest) has officially a 0% missing data. This is because if a patient has a certain comorbidity (e.g., renal insufficiency), the abstractor **actively** checks that variable on the online data submission screen. If it is left blank, the variable is coded as "no" for that condition. This default system could have potential misclassification (e.g., if a patient's condition is not checked despite he or she having that condition).

In a prior audit conducted by the American Heart Association, they found that error rates were about 2.4% of all variables in the registry. Error rates included checking a condition when the patient did not have it, and not checking a condition when the patient did have the condition. (citation for error rate in audit: Peberdy MA, Kaye W, Ornato JP, et al. Cardiopulmonary resuscitation of adults in the hospital: a report of 14720 cardiac arrests from the National Registry of Cardiopulmonary Resuscitation. *Resuscitation.*2003;58:297-308.)

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.7 Disclaimers:

Ad.8 Additional Information/Comments: