**National Quality Forum—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](http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx). * 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**](#Note10) 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**](#Note11) 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**](#Note12)  **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**](#Note13)  **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**](#Note14)**,**[**15**](#Note15) 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**](#Note16) **differences in performance**;  **OR**  there is evidence of overall less-than-optimal performance.  **2b5.** **If multiple data sources/methods are specified, there is demonstration they produce comparable results**.  **2b6.** Analyses identify the extent and distribution of **missing data** (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.  **Notes**  **10.** Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).  **11.** Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.  **12.** Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.  **13.** Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.  **14.** Risk factors that influence outcomes should not be specified as exclusions.  **15.** With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers. |

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

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

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

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

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

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.

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

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

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

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

The 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%) |
| **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.

**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** | | | |  | **P** |
| **Least Black** |  |  | **Most Black** | **All Hospitals n = 288** |
| **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:

Reliability = (hospital-to-hospital variance) ((hospital-to-hospital variance) + (hospital-specific variance))

Between-site (or hospital-to-hospital) variance = αβ

(α + β + 1)(α+β)2

Within-site (or hospital-specific) variance = ί(1 – ί)

nίs

Where,

ί = the proportion of patients who survived to discharge at hospital ί.

nίs = the total number of cardiac arrest events at hospital ί.

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***](#section2b4)

**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***](#section2b5)***.***

**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** Click here to enter number of categories **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 R2 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***](#question2b49)

**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 R2 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 R2 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 R2 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 R2 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**



**2013 DATA**



**2014 DATA**



**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**

P:\ftang\Paul\Standardized risk report\adult\forest 2012 20140703.wmf

**2013 Forest Plot**

P:\ftang\Paul\Standardized risk report\adult\forest 2013 20140703.wmf

**2014 Forest Plot**

P:\ftang\Paul\GWTG\Standardized risk report\Adult 2014\Adult 20150528.wmf

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**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.