**National Quality Forum—Composite Measure Testing (subcriteria 2a2, 2b1-2b6)**

**Measure Number** (*if previously endorsed*)**:** 2377

**Composite Measure Title**: Defect Free Care for AMI

**Date of Submission**: 8/1/20188/1/2018

**Composite Construction:**

Two or more individual performance measure scores combined into one score

All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

|  |
| --- |
| **Instructions: Please contact NQF staff before you begin.**   * If a component measure is submitted as an individual performance measure, the non-composite measure testing form must also be completed and attached to the individual measure submission. * Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form. * **Sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.** * **For composites with 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) and composites (2c) 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. and the 2017 Measure Evaluation Criteria and Guidance. |

|  |
| --- |
| **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 decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [**13**](#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 nonresponses) and how the specified handling of missing data minimizes bias.  **2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:**  **2c1.** the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and  **2c2**.the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.  (*if not conducted or results not adequate, justification must be submitted and accepted)*  **Notes**  **10.** Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).  **11.** Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. 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 different components in the composite, indicate the component after the checkbox. 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: American Hospital Association | other: American Hospital Association |

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

We used a clinical registry, namely the National Cardiovascular Data Registry for Chest Pain- MI Registry, formerly known as the ACTION Registry. This is a national quality improvement registry in which over 1,200 US hospitals participate. Some states and healthcare systems mandate participation. Rigorous quality standards are applied to the data and both quarterly and ad hoc performance reports are generated for participating centers to track and improve their performance.

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

Two different data sets were used to provide support for the testing of reliability and validity. Discharges between January 2016 to December 2017 were used, except where audit data is noted. Hospital information about the proportion of patients with a primary payer source of Medicaid were derived from American Hospital Association 2010 data.

Audit data used was from discharges between Jan 1, 2010 to December 31, 2010. Since measure specifications have not changed, updated audit results have not been provided. The previous audit testing results from the last submission have been kept in this application for reference purposes.

**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*)

The overall measured entities, are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1: Entities Evaluated by Level of Analysis** | | | |
| **Level of Analysis** | **Variable** | **Data Source** | **Number** |
| Patient | Patient Hospital Stay | NCDR Chest Pain – MI Registry | 313,087 |
| Hospital | Facilities | NCDR Chest Pain – MI Registry | 781 |

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

For all the descriptive statistics, we used data collected by the Chest Pain-MI Registry (formerly the ACTION Registry) between January 2016 and December 2017. Descriptive statistics about the patients included in this dataset are provided below (Table 2):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2: Selected Characteristics by Calendar Year** | | | | | | | |
|  |  |  |  |  |  |  |  |
| **Description** | **Total** | | **Year** | | | | **P** |
| **2016** | | **2017** | |
| **#** | **%** | **#** | **%** | **#** | **%** |
|  |  |  |  |  |  |  |  |
| ALL | 313,087 | 100.00 | 157,013 | 100.00 | 156,074 | 100.00 |  |
| **Age>65** |  |  |  |  |  |  | 0.0000 |
| No | 159,829 | 51.05 | 80,726 | 51.41 | 79,103 | 50.68 |  |
| Yes | 153,258 | 48.95 | 76,287 | 48.59 | 76,971 | 49.32 |  |
| **Sex** |  |  |  |  |  |  | 0.4239 |
| Male | 208,133 | 66.48 | 104,273 | 66.41 | 103,860 | 66.55 |  |
| Female | 104,954 | 33.52 | 52,740 | 33.59 | 52,214 | 33.45 |  |
| **Race** |  |  |  |  |  |  | 0.0000 |
| Hispanic | 20,489 | 6.54 | 9934 | 6.33 | 10,555 | 6.76 |  |
| White Non-Hispanic | 246,600 | 78.76 | 124,250 | 79.13 | 122,350 | 78.39 |  |
| Black Non-Hispanic | 34,906 | 11.15 | 17,317 | 11.03 | 17,589 | 11.27 |  |
| Other | 11,092 | 3.54 | 5,512 | 3.51 | 5,580 | 3.58 |  |
| **Insurance** |  |  |  |  |  |  | 0.0000 |
| Medicare | 155,591 | 49.70 | 77,650 | 49.45 | 77,941 | 49.94 |  |
| Medicaid or not private | 27,251 | 8.70 | 13,412 | 8.54 | 13,839 | 8.87 |  |
| Private | 103,414 | 33.03 | 52,575 | 33.48 | 50,839 | 32.57 |  |
| None | 26,831 | 8.57 | 13,376 | 8.52 | 13,455 | 8.62 |  |
| **Composite Measure Performance** |  |  |  |  |  |  | 0.0000 |
| Not Meeting DFC\* | 93,048 | 29.72 | 47,518 | 30.26 | 45,530 | 29.17 |  |
| Meeting DFC\* | 220,039 | 70.28 | 109,495 | 69.74 | 110,544 | 70.83 |  |

\*Abbreviations: DFC, defect free care

**Audit Program:**

To assess inter-rater reliability of the extracted data elements that comprise this measure, we reviewed 330 patients.

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

The datasets, dates, number of measured entities, and number of admissions for all forms of reliability and validity testing were from an uninterrupted 2-year period: 01/2016-12/2017.

**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.

We attributed social risk factors at the hospital-level for the purposes of this analysis. We used Medicaid insurance status as an economic indicator of social risk. We also examined race/ethnicity, age, and gender to determine if there were differences in these demographic indicators of social risk.

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

**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*)

***Note****: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*  
 **Performance measure score** (e.g., *signal-to-noise analysis*)  
  
**2a2.2. Describe the method of reliability testing and what it tests** (*describe the steps―do not just name a method; what type of error does it test; what statistical analysis was used*)

**Split Sample Methodology**

For the performance rates and social risk data, raw rates were calculated and a Pearson correlation coefficient was computed.

**Assessment of item-level reliability through the Audit Program:**

The NCDR Data Quality Program ensures that data submitted to the NCDR are validly collected. The NCDR Data Quality Program consists of 3 main components: data completeness, consistency, and accuracy. Completeness focuses on the proportion of missing data within fields, whereas consistency determines the extent to which logically related fields contain values consistent with other fields. Accuracy characterizes the agreement between registry data and the contents of original charts from the hospitals submitting data. Before entering the Enterprise Data Warehouse (EDW), all submissions are scored for file integrity and data completeness, receiving 1 of 3 scores that are transmitted back to facilities using a color coding scheme. A “red light” means that a submission has failed because of file integrity problems such as excessive missing data and internally inconsistent data. Such data are not processed or loaded into the EDW. A “yellow light” status means that a submission has passed the integrity checks but failed in completeness according to predetermined thresholds. Such data are processed and loaded into the EDW but are not included in any registry aggregate computations until corrected. Facilities are notified about data submission problems and provided an opportunity to resubmit data. Finally, a “green light” means that a submission has passed all integrity and quality checks. Such submissions are loaded to the EDW. After passing the Data Quality Review, data are loaded into a common EDW that houses data from all registries and included for all registry aggregate computations. In a secondary transaction process, data are loaded into registry-specific, dimensionally modeled data marts. A summary of the Program is noted below.

**Data Quality Program Overview**

To assess the reliability of the submitted data in the original NQF application, we provided results of chart audits through which the ACCF was able to compare the reported data with those from an independent assessment by a trained abstractor. Kappas and percentage agreement rates were calculated (referenced below for ease of access).

The National Cardiovascular Data Registry® (NCDR®) Audit Program’s overall purpose is to ensure that data submitted to the NCDR registries are complete, valid, and accurately interpreted and collected, so as to improve the quality of the NCDR registries. Our National Audit program also ensures that benchmarks and comparisons provided to all NCDR participants accurately reflect registry data. Kappa and percentage agreement scores will computed. A summary of the Audit Program is noted below.

|  |  |
| --- | --- |
| **Methodology** | * Nationwide program (i.e., all submitting participants in the United States) * Review of data submitted the previous year * Review of a subset of data elements that can rotate each year * Remote review of data combined with couple of onsite visit * Onsite visits are targeted based on the Data Outlier Program * Random selection of sites and records * Blinded data abstraction from medical charts * Inter-rater Reliability Assessment conducted to validate the audit findings * Adjudication step for participant to refute audit findings |
| **Scope** | * Review of hospital’s medical records for related episodes of care * Assessment of complete submission (Comparison of two lists : hospital list of cases with specific billing codes versus NCDR submitted records) |
| **Criteria for selecting sites/records** | Remote audit :   * Sites passing their quarterly DQR for 2 quarters within audited year * Sites submitting at least the number of records/sites being reviewed   Onsite audit   * Sites identified with an outlier and not contacted with the data outlier program |
| **Scoring** | NCDR uses a grading system for identifying the amount of agreement or matching between the data captured during the medical record review and data submitted to the NCDR. Below are a few definitions for computing these scores and others depending on the Registry:   * The accuracy score for each variable represents how often the NCDR data matches the auditors’ data. * The overall accuracy score is the total number of agreements divided by total number of possible agreements. * Each measurement is summed up for each participant and an overall audit accuracy score is computed. * An overall compliance score is calculated based on the number of matches across two lists. |
|  | *Onsite and Overall accuracy Score*   * A : Above 93% * B : Between 80 and 93% * C : Between 70 and 79% * D : Between 60 and 69% * E : Below 60% |
|  |

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

Split Sample Methodology

The split samples were calculated during the same timeframe to mitigate confounding factors based on time differences. The cohort was split into two random samples to compare measure scores. The distribution of hospital performance was similar in the two samples, and there was an extremely high correlation between hospital performance assessed in the two samples (Pearson correlation coefficient: 0.97179) (Tables 4 and 5).

**Table 4: Split Sample Composition (i.e. Number/proportion of Patients in each sample by Year)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Description** | **Total** | | **Year** | | | | **P-value** |
| **2016** | | **2017** | |
| **#** | **%** | **#** | **%** | **#** | **%** |
|  |  |  |  |  |  |  |  |
| **Random Splitting Samples** |  |  |  |  |  |  | 0.7891 |
| First | 156,659 | 50.04 | 78,527 | 50.01 | 78,132 | 50.06 |  |
| Second | 156,428 | 49.96 | 78,486 | 49.99 | 77,942 | 49.94 |  |

**Table 5: Distribution of Performance for the Defect Free Care Measure Stratified by the Randomly Split Samples**

|  |  |  |
| --- | --- | --- |
| **Description** | **Randomly Split Samples** | |
| **First (RAND=1)** | **Second (RAND=0)** |
|  |  |  |
| N | 781 | 780 |
| Mean | 63.96% | 63.74% |
| Std Deviation | 25.40% | 25.93% |
|  |  |  |
| 100% Max | 100.00% | 100.00% |
| 99% | 96.31% | 96.30% |
| 95% | 92.11% | 92.82% |
| 90% | 89.17% | 89.73% |
| 75% Q3 | 83.74% | 83.23% |
| 50% Median | 71.93% | 72.31% |
| 25% Q1 | 50.00% | 49.55% |
| 10% | 19.35% | 19.29% |
| 5% | 8.93% | 6.95% |
| 1% | 0.00% | 0.00% |
| 0% Min | 0.00% | 0.00% |

Pearson Correlation Coefficient: 0.97179

**Figure 1. Distribution of Performance for the Defect Free Care Measure Stratified by Randomly Split Samples (Top) and by Split Sample Correlation (Bottom)**



Assessment of item-level reliability through Audit Program

NCDR’s Data Quality Program rotates the review of all the variables in the registry. The ACTION registry has 157 elements of which a subset gets reviewed on a 3 year rotating cycle.

**Audit**:

| **CE #** | **Field Name** | **Kappa** | **95% CI** | **Nlevels** | **Final Agreement Rate** |
| --- | --- | --- | --- | --- | --- |
| 3110 | Transferred From Outside Facility | 0.965 | 0.931 - 0.999 | 2 | 99.0% |
| 4030 | STEMI or STEMI Equivalent | 0.878 | 0.827 - 0.93 | 2 | 95.3% |
| 4040 | ECG Findings | 0.777 | 0.713 - 0.842 | 4 | 90.7% |
| 4041 | STEMI or STEMI Equivalent First Noted | 0.733 | 0.665 - 0.8 | 3 | 87.3% |
| 5020 | Current/Recent Smoker (w/in 1 year) | 0.863 | 0.806 - 0.92 | 2 | 95.0% |
| 5070 | Diabetes Mellitus | 0.909 | 0.86 - 0.957 | 2 | 96.3% |
| 6010 | Aspirin in First 24 Hours | 0.384 | 0.139 - 0.63 | 4 | 95.3% |
| 6020 | Aspirin at Discharge | 0.883 | 0.822 - 0.943 | 4 | 96.3% |
| 6270 | Beta Blocker at Discharge | 0.824 | 0.761 - 0.887 | 4 | 94.0% |
| 6320 | ACE Inhibitor at Discharge | 0.811 | 0.758 - 0.864 | 4 | 89.0% |
| 6470 | Statin at Discharge | 0.843 | 0.782 - 0.904 | 4 | 94.3% |
| 7011 | LVEF Not Assessed | 0.602 | 0.469 - 0.736 | 2 | 93.0% |
| 7100 | PCI | 0.987 | 0.97 - 1 | 2 | 99.3% |
| 8000 | Reperfusion Candidate | 0.606 | 0.534 - 0.678 | 3 | 80.7% |
| 8020 | Thrombolytics | 0.669 | 0.592 - 0.746 | 3 | 85.3% |
| 11100 | Discharge Status | 0.934 | 0.844 - 1 | 2 | 99.7% |
| 11101 | Smoking Counseling | 0.756 | 0.688 - 0.823 | 3 | 89.7% |
| 11104 | Cardiac Rehabilitation Referral | 0.386 | 0.314 - 0.458 | 4 | 65.3% |
| 11105 | Discharge Location | 0.808 | 0.746 - 0.87 | 8 | 91.7% |

**2a2.4 What is your interpretation of the results in terms of demonstrating reliability**? (i*.e., what do the results mean and what are the norms for the test conducted?*)

The results of the split sample show a similar performance of the composite measure at discharge for both samples, which demonstrates this is a very reliable measure with an extremely high correlation between hospital performance assessed in the two samples (r = 0.97179).

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

**2b1. VALIDITY TESTING**

***Note****: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

**2b1.1. What level of validity testing was conducted**?

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

**Composite performance measure score**

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

**Validity testing for component measures** *(check all that apply)*

***Note****: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

**Endorsed (or submitted) as individual performance measures**

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

**Empirical validity testing of the component measure score(s)**

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

**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: (Initial testing of this measure):**

NCDR Strategic Quality and Oversight Committee— an ACC leadership oversight committee that serves as the primary resource for crosscutting scientific and quality of care methodological issues – ensured the data dictionaries and metrics are consistent across registries. They also reviewed and approved the methodology and results of the bleeding outcome and model.

These members include Dr. Frederick Masoudi (chair), Dr. David Malenka, Dr. Thomas Tsai, Dr. Matthew Reynolds, Dr. David Shahian, Dr. John Windle, Dr. Fred Resnic, Dr. John Moore, Dr. Deepak Bhatt, Dr. James Tcheng, Dr. Jeptha Curtis, Dr. Paul Chan, Dr. Matt Roe, and Dr. John Rumsfeld.

NCDR Clinical Sub Workgroup is a designated set of experts that oversees this NQF application. Prior to submission, it ensures there is variation in care, disparities data, and that the measure is a true reflection of quality care at a particular site and can also be used to improve quality.

Dr. Jeptha Curtis (chair), Dr. Frederick Masoudi, Dr. John Rumsfeld, Dr. James Jollis and Dr. Deepak Bhatt

NCDR Registry Steering Committee provides strategic direction for the Registry and ensures the measures submitted to NQF met key criterion such as reliability, feasibility, and that there is compelling evidence base behind the development and implementation of this measure. These members include James G. Jollis, M.D., F.A.C.C., Deepak L. Bhatt, M.D., M.P.H., F.A.C.C., Robert L. McNamara, M.D., M.H.S., F.A.C.C., Ivan Rokos, M.D., F.A.C.C., Michael A. Ross, M.D., Michael C. Kontos, M.D., F.A.C.C., Steven V. Manoukian, M.D., F.A.C.C., Harper Stone, MD, Harold L. Dauerman, M.D., F.A.C.C., Martha J. Radford, M.D., F.A.C.C., James A. de Lemos, M.D., F.A.C.C.

Tracy Wang, M.D., F.A.C.C.

Lastly the 16 member NCDR Management Board and 31member ACCF Board of Trustees approved these measures for submission to NQF. The face/content validity of this measure has been achieved by virtue of the noted expertise of those individuals who developed this measure.

**Empirical Validity (Re-endorsement Testing):**

For re-endorsement of this measure, additional empirical validity testing was completed. Empirical analysis was tested by determining if hospitals performed similarly on the defect free care measure and 30-day AMI mortality. The testing focused on construct validation which tested the hypothesis that use of defect free care processes for AMI patients may lead to better outcomes. This was achieved by examining the distribution and correlation of the defect free care (DFC) composite score and the 30-day risk-standardized mortality rates (RSMR) for AMI from admission to 30-days. The variables in the model included age, heart rate, systolic blood pressure, troponin ratio, and creatinine level (*McNamara et al. Development of a hospital outcome measure intended for use with electronic health records: 30-day risk-standardized mortality after acute myocardial infarction. Med Care, 2015, vol. 53 (pg. 818-26*). For this specific analysis, the study period included Q4 2013 to Q3 2014 as this comprised the latest NDI-ACTION linked data available at the time of the analysis.

Critical Data Element Validity:

See audit table in question 2a2.3 for critical data elemt results.

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

**Face Validity**

Face validity was achieved through reaching consensus that the measure had strong clinical evidence and was reliable.

**Empirical Validity**

Below are the results achieved from the empirical validity testing (Table 6):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 6: Distribution of Performance Rates for DFC and RSMR in the Time Period 2013Q4 to 2014Q3 (N=781)** | | | | |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Description** | **DFC\*** | | **RSMR\*\*** | | |  | **Performance rate (%)** |  | **Performance rate (%)** | |  |  |  |  |  | | Mean |  | 63.2% |  | 6.3% | | Std Deviation |  | 27.0% |  | 1.1% | |  |  |  |  |  | | 100% Max |  | 100.0% |  | 11.5% | | 99% |  | 96.8% |  | 9.3% | | 95% |  | 93.7% |  | 8.3% | | 90% |  | 90.5% |  | 7.9% | | 75% Q3 |  | 84.3% |  | 6.9% | | 50% Median |  | 72.5% |  | 6.2% | | 25% Q1 |  | 47.5% |  | 5.5% | | 10% |  | 16.1% |  | 4.9% | | 5% |  | 5.9% |  | 4.6% | | 1% |  | 0.0% |  | 4.0% | | 0% Min |  | 0.0% |  | 3.5% | |  |  |  |  |

**Pearson correlation coefficient between DFC and RSMR -0.1093**

*\*DFC = Defect Free Care*

*\*\*RSMR = Risk Standardize Mortality Rate*

**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?*)

**Face validity**: The individual components have been associated with better outcomes and are accepted quality measures in patient populations. As noted in Section 2a2.4., we have good evidence of validity from the Chest Pain – MI Registry audit data. A vast majority of data elements showing >90% agreement between routinely collected data and audit data.

**Empirical validity:** The median rate of delivering defect free care was 72.5% (IQR: 47.5% to 84.3%), and the median mortality rate at 30 days was 6.2% (IQR: 5.5% to 6.9%). There was a similar distribution of hospitals by volume across both measures. The negative correlation coefficient was significant and in the hypothesized direction, such that a higher group of patients receiving defect free care was associated with lower mortality rates. Yet, the correlation is relatively low (-0.11), which is not surprising when comparing a process of care measure to an outcome measure. The low correlation may be explained by the fact that there are a number of other unmeasured factors that could contribute to 30-day mortality rates beyond whether defect free care was delivered in-hospital (e.g., unsuccessful procedure, lack of follow-up, poor medication adherence or access to care). Further, the 30-day time period started upon admission to the hospital thus the rates also accounted for in-hospital mortality. In sum, the empirical validation demonstrates there is a relationship, albeit statistically a small one, between defect free care and short-term mortality.

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

**2b2. EXCLUSIONS ANALYSIS**

***Note:*** *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*

**NA**  **no exclusions — *skip to section*** [***2b4***](#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*)

The exclusions for this measure were minimal and comprised: patients <18 years of age, hospital submissions that did not pass the NCDR quality check, and patients who were ineligible for defect free care measure (e.g., contraindications, clinical studies).

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

Table 6 (below) provides information about how the final sample was derived. “Not eligible for the composite measure” means those patients who are not eligible for any of the 8 components of the STEMI measures or any of the 5 components of the NSTEMI measures. Ineligible reasons include contraindications or those individuals enrolled in clinical trial studies.

|  |
| --- |
|  |
|  | |  |  |
| **Exclusions** | | **Number of Hospital Stay** | **Number of Facilities** |
| **#** | **#** |  |
|  | |  |  |  |
| *Initial Sample* | | 1,398,006 | 1,229 |  |
| Discharges not between 2016 and 2017 | | - 989,668 | 166 |  |
| *Remaining* | | 408,338 | 1,063 |  |
| Age<18 | | - 0 | 0 |  |
| *Remaining* | | 408,338 | 1,063 |  |
| Hospital submission did not pass the data quality  check | | - 82,589 | 282 |  |
| *Remaining* | | 325,749 | 781 |  |
| Not eligible to the defect free care measure | | - 12,662 | 0 |  |
| **Study Sample** | | **313,087** |  |

**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*)

There are no discretionary exclusions, and these exclusions only pertain to variables that are necessary to derive a precise measure of quality. It would not be advisable to include information from hospitals that did not meet quality thresholds or include patients who are not eligible for individual components of the composite measure.   
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES**  
***Note:***  *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*  
***If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section*** [***2b4***](#section2b5)***.***

**2b3.1. What method of controlling for differences in case mix is used?** *(check all that apply)*

**Endorsed (or submitted) as individual performance measures**

**No risk adjustment or stratification**

**Statistical risk model with** Click here to enter number of factors **risk factors**

**Stratification by** Click here to enter number of categories **risk categories**

**Other,** Click here to enter description

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

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

**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?

N/A

**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?**

N/A

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

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

*Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below*.  
***If stratified, skip to*** [***2b3.9***](#question2b49)

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

N/A

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

N/A

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

N/A

**2b3.9. Results of Risk Stratification Analysis**:

N/A

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

N/A

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

N/A

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

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

***Note:*** *Applies to the composite performance measure.*

**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)*

We examined variation in hospital performance for the composite measure based on overall performance, and stratified by subgroups of sex, age, race, and the proportion of patients who are insured through Medicaid to identify if there were meaningful differences in social risk.

**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*)

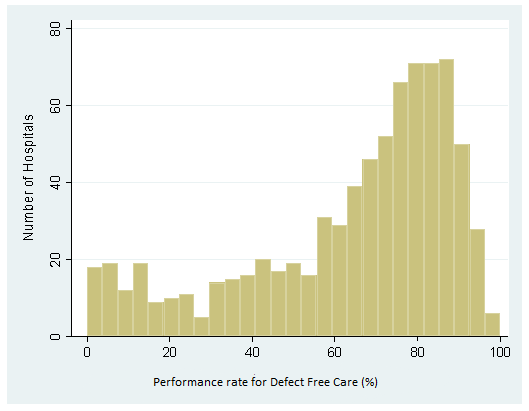
**Overall**

The median rate of performance for defect free care across all hospitals was 71.7%. There was considerable variation in providing defect free care, ranging from 50.1% to 83.2% for the first and third quartiles of hospitals, respectively (Table 7), and the distribution was right-skewed such that the majority of hospitals, between 60% to 100%, provided defect free care (Figure 1).

**Table 7: Distribution of Performance of Defect Free Care**

|  |  |  |
| --- | --- | --- |
| Description |  | DFC (%) |
|  |  |  |
| Mean |  | 63.87 |
| Std Deviation |  | 25.50 |
|  |  |  |
| 100% Max |  | 100.00 |
| 99% |  | 95.86 |
| 95% |  | 92.22 |
| 90% |  | 89.29 |
| 75% Q3 |  | 83.25 |
| 50% Median |  | 71.70 |
| 25% Q1 |  | 50.14 |
| 10% |  | 19.07 |
| 5% |  | 7.75 |
| 1% |  | 0.00 |
| 0% Min |  | 0.00 |
|  |  |  |

**Figure 1: Histogram of Performance of Defect Free Care**



**Subgroups**

Across stratified analyses based on sex, age, race, and proportion of patients who are insured through Medicaid, we found significant overlap in the distribution of hospital performance, as detailed below.

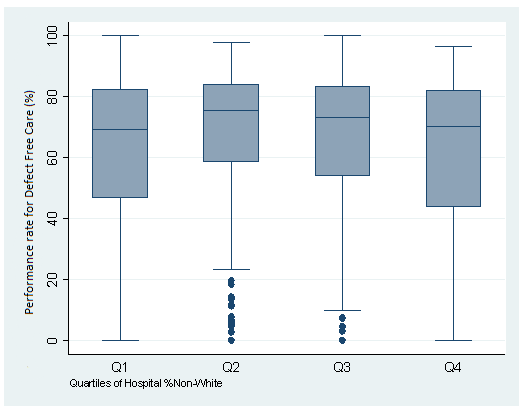
***Proportion of Non-White***

Hospitals (n=781) were stratified into quartiles by their proportion of non-White patients (median: 10.4%, IQR: 4.6% to 22.3%). Hospital performance across quartiles was similar regardless of the percent of non-White patients hospitals treated, with median performance ranging from 69.1% to 75.5% (Table 9; Figure 2).

**Table 9: Distribution of Performance Rates for the Defect Free Care Measure Stratified by Hospital Quartiles of Non-White Patients (N=781)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Description** | **Non-White (%)** | | | |
| **Q1** | **Q2** | **Q3** | **Q4** |
|  |  |  |  |  |
| N | 195 | 195 | 196 | 195 |
| Mean | 62.6% | 67.0% | 65.4% | 60.5% |
| Std Deviation | 25.8% | 24.6% | 24.0% | 27.1% |
|  |  |  |  |  |
| 100% Max | 100.0% | 97.8% | 100.0% | 96.6% |
| 99% | 97.6% | 95.3% | 96.8% | 94.6% |
| 95% | 92.1% | 91.7% | 92.6% | 92.4% |
| 90% | 88.7% | 89.4% | 90.0% | 89.3% |
| 75% Q3 | 82.4% | 84.2% | 83.6% | 82.1% |
| 50% Median | 69.1% | 75.5% | 73.2% | 70.0% |
| 25% Q1 | 46.8% | 58.6% | 54.0% | 43.8% |
| 10% | 19.1% | 23.1% | 24.5% | 13.7% |
| 5% | 5.7% | 7.5% | 11.4% | 5.5% |
| 1% | 0.0% | 0.0% | 3.0% | 0.0% |
| 0% Min | 0.0% | 0.0% | 0.0% | 0.0% |

**Figure 2: Distribution of Performance Rates of the Defect Free Care Measure Stratified by Quartiles of Non-White Patients at the Hospital-Level**



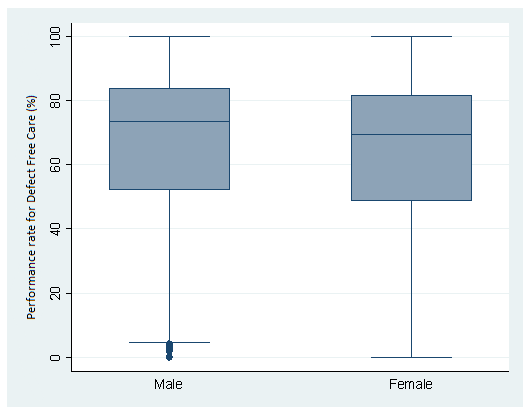
***Gender***

The median hospital performance among female patients was 69.5% (IQR; 48.6% to 81.8%) while among male patients it was slightly higher at 73.5% (IQR: 52.1% to 83.9%) (Table 10 and Figure 3).

**Table 10: Distribution of the Performance Rates for the Defect Free Care Measure Stratified by Gender at the Hospital-Level (N=781)**

|  |  |  |
| --- | --- | --- |
| **Description** |  | |
| **Female** | **Male** |
|  |  |  |
| N | 779 | 781 |
| Mean | 62.2% | 64.8% |
| Std Deviation | 25.5% | 25.8% |
|  |  |  |
| 100% Max | 100.0% | 100.0% |
| 99% | 96.4% | 96.3% |
| 95% | 91.0% | 92.9% |
| 90% | 88.4% | 90.1% |
| 75% Q3 | 81.8% | 83.9% |
| 50% Median | 69.5% | 73.5% |
| 25% Q1 | 48.6% | 52.1% |
| 10% | 19.2% | 19.7% |
| 5% | 7.6% | 8.7% |
| 1% | 0.0% | 0.0% |
| 0% Min | 0.0% | 0.0% |

**Figure 3: Distribution of the Performance Rates of the Defect Free Care Measure Stratified by Gender at the Hospital-Level**

****

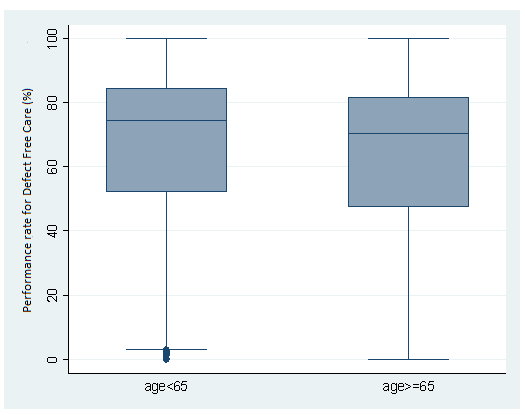
***Age***

The median hospital performance in delivering Defect Free Care among patients aged less than 65 years was 74.6% (IQR: 52.0% to 84.7%) while that among patients aged 65 years or greater was 70.5% (IQR: 47.3% to 81.7%) (Table 11 and Figure 4).

**Table 11: Distribution of Performance Rates of the Defect Free Care Measure Stratified by Age at the Hospital-level (N=781)**

|  |  |  |
| --- | --- | --- |
| **Description** | **Age > 65** | |
| **Yes** | **No** |
|  |  |  |
| Mean | 62.5% | 65.2% |
| Std Deviation | 25.1% | 26.4% |
|  |  |  |
| 100% Max | 100.0% | 100.0% |
| 99% | 95.7% | 97.3% |
| 95% | 91.5% | 93.5% |
| 90% | 88.3% | 90.8% |
| 75% Q3 | 81.7% | 84.7% |
| 50% Median | 70.5% | 74.6% |
| 25% Q1 | 47.3% | 52.0% |
| 10% | 20.0% | 17.5% |
| 5% | 9.1% | 6.9% |
| 1% | 0.0% | 0.0% |
| 0% Min | 0.0% | 0.0% |

**Figure 4: Distribution of the Performance Rates of the Defect Free Care Measure Stratified by Age Group at the Hospital-Level**

****

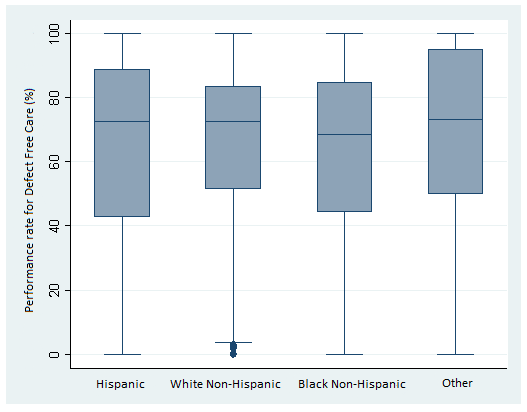
***Race/Ethnicity***

The distribution of hospital performance was examined among White (non-Hispanic), Black (non-Hispanic), Hispanic and Other race patients. There was significant overlap in hospital performance with median performance ranging from 68.4% for Black non-Hispanic to 73.3% for other race. White non-Hispanic and Hispanic had the same median performance (72.7%) (Table 12 and Figure 5).

**Table 12: Distribution of the Performance of the Defect Free Care Measure Stratified by Race at the Hospital-Level (N=781)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Description** | **Race** | | | |
| **Hispanic** | **White non-Hispanic** | **Black non-Hispanic** | **Other** |
|  |  |  |  |  |
| Mean | 63.1% | 64.4% | 62.1% | 65.7% |
| Std Deviation | 31.8% | 25.6% | 29.5% | 32.3% |
|  |  |  |  |  |
| 100% Max | 100.0% | 100.0% | 100.0% | 100.0% |
| 99% | 100.0% | 97.6% | 100.0% | 100.0% |
| 95% | 100.0% | 92.8% | 100.0% | 100.0% |
| 90% | 100.0% | 89.2% | 100.0% | 100.0% |
| 75% Q3 | 88.9% | 83.6% | 85.0% | 95.1% |
| 50% Median | 72.7% | 72.7% | 68.4% | 73.3% |
| 25% Q1 | 42.9% | 51.4% | 44.4% | 50.0% |
| 10% | 7.6% | 20.5% | 12.5% | 0.0% |
| 5% | 0.0% | 8.3% | 0.0% | 0.0% |
| 1% | 0.0% | 0.0% | 0.0% | 0.0% |
| 0% Min | 0.0% | 0.0% | 0.0% | 0.0% |

**Figure 5: Distribution of the Performance Rates of the Defect Free Care Measure Stratified by Race/Ethnicity at the Hospital-Level**

****

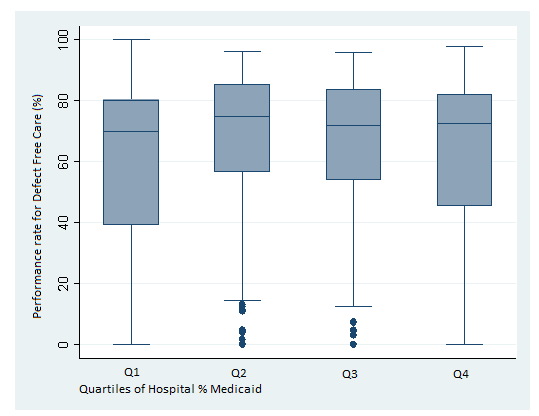
***Insurance***

Hospitals (n=781) were stratified into quartiles by their proportion of patients with Medicaid as the primary insurance (median: 9.9%, IQR: 6.0% to 15.1%). Hospital performance was similar across hospitals stratified into quartile by the proportion of patients they care for who have Medicaid insurance coverage. Median hospital performance ranged from 69.1% (Quartile 1) to 75.5% (Quartile 2) (Table 13 and Figure 6).

**Table 13: Distribution of Performance Rates of the Defect Free Care Measure Stratified by Quartile of Hospital Percent Medicaid (N=781)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Description** | **Medicaid** | | | |
| **Q1** | **Q2** | **Q3** | **Q4** |
|  |  |  |  |  |
| N | 195 | 195 | 196 | 195 |
| Mean | 62.6% | 67.0% | 65.4% | 60.5% |
| Std Deviation | 25.8% | 24.6% | 24.0% | 27.1% |
|  |  |  |  |  |
| 100% Max | 100.0% | 97.8% | 100.0% | 96.6% |
| 99% | 97.6% | 95.3% | 96.8% | 94.6% |
| 95% | 92.1% | 91.7% | 92.6% | 92.4% |
| 90% | 88.7% | 89.4% | 90.0% | 89.3% |
| 75% Q3 | 82.4% | 84.2% | 83.6% | 82.1% |
| 50% Median | 69.1% | 75.5% | 73.2% | 70.0% |
| 25% Q1 | 46.8% | 58.6% | 54.0% | 43.8% |
| 10% | 19.1% | 23.1% | 24.5% | 13.7% |
| 5% | 5.7% | 7.5% | 11.4% | 5.5% |
| 1% | 0.0% | 0.0% | 3.0% | 0.0% |
| 0% Min | 0.0% | 0.0% | 0.0% | 0.0% |

**Figure 6: Distribution of Performance Rates of the Defect Free Care Measure Stratified by Quartile of Hospital Percent Medicaid**

****

**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 wide gap in performance rates, along with broad interquartile ranges, across various stratified populations demonstrates that this measure is necessary to improve the quality gap.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

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

***Note:***  *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*

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

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

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

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

N/A

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

N/A  
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

***Note:***  *Applies to the overall composite measure.*

**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*)

There were no missing data for this measure. Any hospitals with missing data were excluded from the measure as they would not have passed the NCDR data quality review.

**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*)

N/A

**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*)

N/A

**2c. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH**

***Note:*** *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

**2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.**

We believe the face/content validity of this measure has been achieved by virtue of the noted expertise of those individuals who developed this measure. The individual components of the composite have already shown to impact clinical outcomes.

The empirical validity analysis demonstrated that the individual component measures fit the overall quality construct by assessing the correlation of the defect free care measure with its components, including: Aspirin at Arrival, Aspirin prescribed at Discharge, Beta-Blocker Prescribed at Discharge, Statin Prescribed at Discharge, Evaluation of LV Systolic Function, ACEI or ARB for LVSD at Discharge, Time to Fibrinolytic Therapy: <30 minutes, Time to Balloon: <90 minutes, Reperfusion Therapy, Adult Smoking Cessation Advice Counseling, and Cardiac Rehabilitation Patient Referral From an Inpatient Setting.

**2d1.1 Describe the method used** (*describe the steps―do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

We computed hospital-level measures for the eleven measure components individually and then correlated the results with the hospital-level composite results using Pearson correlation.

**2d1.2. What were the statistical results obtained from the analysis of the components?** (e.g., *correlations, contribution of each component to the composite score, etc*.; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

The Pearson correlation coefficients between the defect free care measure and its components are available in Tables 13 and 14.

**Table 13: Distribution of Overall Defect Free Care and its Components at the Hospital-Level (N=781)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Description** | **DFC Composite**  **Value** | **Aspirin at Arrival**  **Value** | **Aspirin at Discharge**  **Value** | **Beta Blocker at Discharge**  **Value** |
|  |  |  |  |  |
| Mean | 63.9% | 98.0% | 96.9% | 95.1% |
| Std Deviation | 25.5% | 3.3% | 5.2% | 7.2% |
|  |  |  |  |  |
| 100% Max | 100.0% | 100.0% | 100.0% | 100.0% |
| 99% | 95.9% | 100.0% | 100.0% | 100.0% |
| 95% | 92.2% | 100.0% | 100.0% | 100.0% |
| 90% | 89.3% | 100.0% | 100.0% | 99.5% |
| 75% Q3 | 83.3% | 100.0% | 99.2% | 98.7% |
| 50% Median | 71.7% | 99.2% | 98.2% | 97.4% |
| 25% Q1 | 50.1% | 97.6% | 96.3% | 94.0% |
| 10% | 19.1% | 94.3% | 93.0% | 88.9% |
| 5% | 7.8% | 92.1% | 90.0% | 84.8% |
| 1% | 0.0% | 85.6% | 76.2% | 68.9% |
| 0% Min | 0.0% | 66.7% | 42.9% | 0.0% |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Description** | **Statin at Discharge**  **Value** | | **LV Systolic Function Value** | | | | **ACE/ARB LVSD**  **Value** | **Fibrinolytic <30m**  **Value** | |
|  |  | |  | | | |  |  | |  | |  |  |
| Mean | 98.5% | | 95.0% | | | 87.5% | | 44.7% | | |
| Std Deviation | 1.9% | | 6.3% | | | 13.9% | | 44.4% | | |
|  |  | |  | | |  | |  | | |
| 100% Max | 100.0% | | 100.0% | | | 100.0% | | 100.0% | | |
| 99% | 100.0% | | 100.0% | | | 100.0% | | 100.0% | | |
| 95% | 100.0% | | 100.0% | | | 100.0% | | 100.0% | | |
| 90% | 100.0% | | 99.4% | | | 100.0% | | 100.0% | | |
| 75% Q3 | 99.6% | | 98.2% | | | 97.3% | | 100.0% | | |
| 50% Median | 99.0% | | 96.5% | | | 91.7% | | 50.0% | | |
| 25% Q1 | 98.0% | | 94.1% | | | 82.9% | | 0.0% | | |
| 10% | 96.4% | | 90.0% | | | 68.3% | | 0.0% | | |
| 5% | 95.1% | | 85.6% | | | 60.0% | | 0.0% | | |
| 1% | 88.9% | | 66.7% | | | 40.0% | | 0.0% | | |
| 0% Min | 83.3% | | 22.2% | | | 0.0% | | 0.0% | | |
|  |  | |  | | |  | |  | | |
| **Description** | | **Fibrinolytic <90m**  **Value** | | **Reperfusion Therapy**  **Value** | **Smoking Cessation**  **Value** | | | | **Cardiac Rehab Referral**  **Value** | | | | |
|  | |  | |  |  | | | |  | | | | |
| Mean | | 94.0% | | 90.2% | 96.7% | | | | 71.3% | | | | |
| Std Deviation | | 6.8% | | 9.5% | 8.0% | | | | 29.7% | | | | |
|  | |  | |  |  | | | |  | | | | |
| 100% Max | | 100.0% | | 100.0% | 100.0% | | | | 100.0% | | | | |
| 99% | | 100.0% | | 100.0% | 100.0% | | | | 100.0% | | | | |
| 95% | | 100.0% | | 100.0% | 100.0% | | | | 99.1% | | | | |
| 90% | | 100.0% | | 98.9% | 100.0% | | | | 97.9% | | | | |
| 75% Q3 | | 98.2% | | 95.9% | 100.0% | | | | 93.8% | | | | |
| 50% Median | | 96.0% | | 92.6% | 100.0% | | | | 83.4% | | | | |
| 25% Q1 | | 92.3% | | 87.3% | 97.1% | | | | 57.3% | | | | |
| 10% | | 86.8% | | 79.9% | 90.5% | | | | 14.1% | | | | |
| 5% | | 81.5% | | 72.7% | 83.3% | | | | 4.1% | | | | |
| 1% | | 66.7% | | 57.0% | 57.9% | | | | 0.0% | | | | |
| 0% Min | | 39.5% | | 0.0% | 0.0% | | | | 0.0% | | | | |

**Table 14: Pearson Correlation Coefficient between Defect Free Care and its Components**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Components** |  | **r** |
|  | Aspirin at Arrival |  | 0.5027 |
|  | Aspirin at Discharge |  | 0.4463 |
|  | BB at Discharge |  | 0.5062 |
|  | Statin at Discharge |  | 0.3408 |
|  | LV Systolic Function |  | 0.3993 |
|  | ACE/ARB LVSD |  | 0.4329 |
|  | Fibrinolytic <30m |  | 0.2261 |
|  | Fibrinolytic <90m |  | 0.4025 |
|  | Reperfusion Therapy |  | 0.1166 |
|  | Smoking Cessation |  | 0.4247 |
|  | Cardiac Rehab Refer |  | 0.9394 |

**2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite?** (i*.e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

The results of the empirical validity testing demonstrate a correlation between all components of the measure and the overall performance of defect free care. Most elements have a moderate correlation to the overall composite measure, with some having a very strong (e.g., cardiac rehabilitation referral) and weaker (e.g., reperfusion therapy) correlations. While there is variation across all components, we feel all components are important aspects in delivering defect free care. The inclusion of all variables is also explained through achieving face validity.

**2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible**

**2d2.1 Describe the method used** (*describe the steps―do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

This is an all-or-none composite, thus no empirical analyses pertinent to aggregations or weighting were conducted. The components mentioned throughout the application are part of the composite measure indicator definition, not the composite of different measures.

**2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules?** (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

N/A

**2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct?** (i*.e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

N/A