



**UHC ANNUAL CONFERENCE 2011**  
September 21–23 • Sheraton Chicago Hotel & Towers • Chicago, Ill

# **UHC 2011 Quality and Accountability Study**

## **Scoring and Ranking Methodology**

**September 2011**

## **Background**

Why do some organizations succeed in consistently providing high-quality care?

In 2005, UHC embarked on a watershed project, the Quality and Accountability Study, to identify structures and practices that are associated with high performance in quality and safety across a wide variety of patient populations. To distinguish high-performing organizations from the rest, a scoring and ranking model for academic medical centers was developed. The 2005 study ultimately identified 5 key attributes of high-performing organizations:

- **A shared sense of purpose**
- **Leadership style**
- **An accountability system**
- **A focus on results**
- **A culture of collaboration**

The full details of the study were published in *Academic Medicine* in December 2007. UHC has conducted many member retreats to explain the methodology used to score and rank performance and, more importantly, to discuss the factors that the study has shown to be critical to organizational success. UHC has conducted additional site visits to validate and refine its understanding of the links between leadership and quality.

UHC continues to refine the scoring and ranking methodology to ensure that contemporary measures of quality and safety are included in the rankings. The quality and accountability system has allowed UHC member organizations to compare their year-over-year performance with that of other academic medical centers and target opportunities for improvement. This document outlines the methodology used in the 2011 Quality and Accountability Study.

## **Methods**

### ***Principles of Structural Development***

As in 2005, the 2011 study was designed to encompass as broad a spectrum of patient activity as possible. The Institute of Medicine's 6 domains of care (safety, timeliness, effectiveness, efficiency, equity, and patient centeredness—commonly referred to as STEEEP) were again used as a guide in structuring the performance categories, and measures of mortality, safety, effectiveness, equity, and patient centeredness were specifically identified and included to determine rankings. For the first time in 2011, metrics for efficiency, previously provided only to give some perspective on performance, were also factored into the overall institutional rankings. Timeliness data, focusing on an organization's ability to schedule and provide services in a timely manner, are embedded in some of the measures of effectiveness. However, more routinely collected measures are needed in this area.

To allow accurate comparisons between organizations, only full UHC members were used to establish normative performance; 101 member institutions were included in the 2011 analysis. Source data were obtained from UHC's Clinical Data Base (CDB), UHC's Core Measures database, and publicly reported Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) data. To reduce variations in quality measures and facilitate understanding of the results, nationally recognized and accepted measure definitions were used as much as possible. UHC's selection and weighting of measures was overseen by members of the UHC Quality and Accountability Study steering committee, as well as members of the Performance Improvement/Comparative Data & Informatics subcommittee of UHC's Board of Directors. (Appendix A lists the members of these groups.)

### ***Data Sources, Domains, and Measures***

Institutional performance metrics were grouped into 6 domains: Mortality, Effectiveness, Safety, Equity, Patient Centeredness, and Efficiency. To evaluate sustained performance and minimize the effect of seasonal fluctuations on the results, a full year of data was used for all domains. For the Mortality, Effectiveness, Safety, and Efficiency domains, CDB and Core Measures data from the most recent year of available data (i.e., Q3 2010 through Q2 2011 for CDB and Q2 2010 through Q1 2011 for Core Measures) were used. Patients with an admit source of transfer from hospice and nonviable neonates were excluded from the mortality, safety, readmission, and efficiency metrics. The most recent HCAHPS data available from the Hospital Compare Web site (Q4 2009 through Q3 2010) were used for the Patient Centeredness domain.

### *Important Data Source Changes for 2011*

In 2011, the CDB completed a significant adjustment of its product line structure, moving from product lines to service lines. While the service lines have names similar to those of the product lines used in the past, the service lines were created to better reflect assignment of trauma patients and to remove the potentially disparate impact of some minor procedures (e.g., interruption of vena cava) on service line assignment.

Additionally, UHC's risk models in the CDB were significantly revised in 2011 to provide a clearer picture of the patient's condition on admission compared with condition at discharge. Revisions included:

- A reduction in the use of risk of mortality (ROM) as a predictor variable in mortality models. Instead, specific conditions and diagnoses that would be used in calculating an ROM value (e.g., sepsis, shock, pneumonia, trauma) were used. The admit ROM is now the predictive variable for cases in which ROM is still used.
- Incorporation of the present on admission (POA) indicator when reviewing diagnoses that are contributing factors to mortality to focus on conditions that were present at the time of admission.

The following sections discuss each domain in more detail; a comprehensive description of metrics used, exclusions, data transformations, and scoring are included in Appendix B of this document.

### *Mortality*

The 2011 Mortality domain was scored using a hybrid approach to account for both system-level and service line-level performance. The rationale for this approach is that factors affecting the risk of death occur at both the service-line level and the system level. The Mortality domain score has 2 equal components: an aggregate mortality score and a composite of individually scored service lines. The aggregate mortality score is the observed-to-expected (O/E) mortality ratio for all patients in 28 UHC service lines; however, because of the overlap of Medicare severity diagnosis-related group (MS-DRG) codes, 2 of the 27—"heart transplant or implant of heart assist system" and "lung transplant"—were combined into a single service line, "heart/lung transplant." Individual cardiac surgery and thoracic surgery service lines were created for 2011, but for the purposes of this study were combined into a single service line, "cardiothoracic surgery." The other 27 service lines scored for the aggregate mortality component are:

- Bone marrow transplant
- Burns
- Cardiology
- Cardiothoracic surgery
- Gastroenterology
- Gynecology
- Gynecologic oncology
- Heart/lung transplant

- HIV
- Kidney/pancreas transplant
- Liver transplant
- Medical oncology
- Medicine, general
- Neurology
- Neurosurgery
- Obstetrics
- Orthopedics
- Otolaryngology
- Plastic surgery
- Rheumatology
- Spinal surgery
- Surgical oncology
- Surgery, general
- Trauma
- Urology
- Vascular surgery
- Ventilator support

The aggregate mortality O/E distributions for all service lines were reviewed and evaluated. Eight service lines with significant mortality continue to be the best indicators of performance across the membership. The 8 service lines selected are those considered to be both large and strategic services for academic medical centers and therefore provide a good opportunity to evaluate performance at the service-line level:

- Cardiology
- Cardiothoracic surgery (includes cardiac surgery and thoracic surgery service lines for 2011)
- Gastroenterology
- Medical oncology
- Medicine, general
- Neurology
- Neurosurgery
- Surgery, general

Observed-to-expected mortality ratios for each of these 8 service lines were determined and weighted equally in calculating the service-line score. The product-line score and the aggregate score contribute equally to the final mortality score. All mortality calculations used UHC's risk-adjustment models, which are based on MS-DRG codes.

### *Effectiveness*

The Effectiveness domain includes an overall hospital readmission rate and Joint Commission Hospital Core Measures performance scores for 4 patient groups: acute myocardial infarction (AMI), heart failure (HF), pneumonia (PN), and Surgical Care Improvement Project (SCIP). The metrics in detail are:

**All-cause 30-day readmission rate:** The percentage of admitted patients who return for any reason within 30 days of discharge. For the 2011 rankings, because of the intense national focus on readmissions as defined by the Centers for Medicare & Medicaid Services (CMS), the steering committee advised using the all-cause readmission rate rather than the related readmission rate used in previous years'

rankings. Expected readmissions for chemotherapy, radiation therapy, routine inpatient dialysis, obstetrics, and rehabilitation were excluded from the numerator of the metric. Readmissions with a major diagnostic category (MDC) code of 19 (mental diseases and disorders) or 20 (alcohol/drug use and alcohol/drug-induced organic mental disorders) who were readmitted within 1 day of discharge were also excluded from the numerator of the readmission metric. It is believed that this restriction will account for patients who are discharged from an acute service and immediately admitted into a psychiatric or substance abuse treatment unit.

**Joint Commission Hospital Core Measures performance score:** The percentage of patients who received all care for which they were eligible, as determined by measure definitions. This model “bundles” all of the required elements of care for each measure set and evaluates hospital performance at the patient level. This represents a more stringent performance requirement than had been used in past reporting by The Joint Commission and CMS, which reviewed hospital performance for each measure rather than assessing each patient’s care according to the standard. The specific measures endorsed by the Hospital Quality Alliance are:

- Acute myocardial infarction composite (AMI-1, AMI-2, AMI-3, AMI-4, AMI-5, AMI-7a, and AMI-8a)
- Heart failure composite (HF-1, HF-2, HF-3, and HF-4)
- Pneumonia composite (PN-2, PN-3a, PN-3b, PN-4, PN-5c, PN-6, and PN-7)
- Surgical Care Improvement Project composite (SCIP-Inf-1a, SCIP-Inf-2a, SCIP-Inf-3a, SCIP-Inf-4, SCIP-Inf-6, SCIP-Inf-9, SCIP Inf-10 SCIP-Card-2, SCIP-VTE-1, and SCIP-VTE-2)

### *Safety*

The Safety domain is based on the Patient Safety Indicators (PSIs) developed by the Agency for Healthcare Research and Quality (AHRQ). The rate distributions among UHC members for the 23 PSIs were reviewed and 6 that showed sufficient variation and incidence were selected for inclusion in the 2011 Safety domain scoring:

- PSI-3: Pressure ulcer
- PSI-6: Iatrogenic pneumothorax
- PSI-7: Central line–associated bloodstream infections
- PSI-9: Postoperative hemorrhage and hematoma
- PSI-11: Postoperative respiratory failure
- PSI-12: Postoperative pulmonary embolism or deep vein thrombosis

Historically, the selection of measures for this domain was based on the metrics' signal ratios as determined by AHRQ. UHC's experience has shown that PSIs with high signal ratios are still subject to high false-positive rates; therefore, all PSIs are evaluated annually for potential inclusion in this domain.

Observed-to-expected rates based on AHRQ's PSI version 4.2 were used for PSI-6, PSI-7, PSI-9, PSI-11, and PSI-12. For PSI-3, the AHRQ PSI version 3.2 observed-to-expected rate was used because version 4.2 risk-adjustment models for this PSI were found to be unreliable.

### *Equity*

Equity was measured using the 4 clinical core measures used for Effectiveness. For each clinical area, hospitals' bundled core measure results were evaluated for statistically significant differences in performance rates in each of 3 dimensions: gender, race (white versus nonwhite), and socioeconomic status (payer classification of Medicaid, county medically indigent, charity, self-pay/uninsured, and Title V maternal/child health versus all other payer types).

### *Patient Centeredness*

In past iterations of the UHC Quality and Accountability Study, question 21 on the HCAHPS survey—"Using any number from 0 to 10, where 0 is the worst hospital possible and 10 is the best hospital possible, what number would you use to rate this hospital during your stay?"—has served as the source for the Patient Centeredness metric. In 2011, the steering committee recommended inclusion of the questions assigned to the CMS categories noted below:

- **Category:** Nurse Communication
  - How often did nurses communicate well with patients?
- **Category:** Doctor Communication
  - How often did doctors communicate well with patients?
- **Category:** Pain Management
  - How often was patients' pain well controlled?
- **Category:** Communication About Medications
  - How often did staff explain about medicines before giving them to patients?
- **Category:** Cleanliness and Quietness
  - How often was the area around patients' rooms kept quiet at night?
  - How often were the patients' rooms and bathrooms kept clean?
- **Category:** Responsiveness of Hospital Staff
  - How often did patients receive help quickly from hospital staff?

- **Category:** Discharge Information
  - Were patients given information about what to do during their recovery at home?
- **Category:** Overall Rating of Hospital
  - How do patients rate the hospital overall?
  - Would patients recommend the hospital to friends and family?

UHC accessed the risk-adjusted data available for download on the Department of Health and Human Services Hospital Compare Web site ([www.hospitalcompare.hhs.gov](http://www.hospitalcompare.hhs.gov)). The most recent 12 months of data available from this site were for the period Q4 2009 through Q3 2010.

### *Efficiency*

Efficiency metrics are focused on per-case performance on cost and LOS from the UHC CDB.

Measures include:

- LOS observed/expected ratio (excluding LOS outliers and early deaths)
- Direct cost observed/expected ratio

The metrics were calculated for the 8 service lines individually evaluated in the Mortality domain. To address potential variations in selection of care venues between inpatient and ambulatory settings across hospitals, discharges with a length of stay of 1 day and one of a set of MS-DRG codes that are areas of focus for Recovery Audit Contractors (RACs) were excluded from the calculations. The MS-DRGs affected are:

- 069: Transient ischemia
- 190: Chronic obstructive pulmonary disease with major complication/comorbidity (CC)
- 191: Chronic obstructive pulmonary disease with CC
- 192: Chronic obstructive pulmonary disease without CC/major CC
- 291: Heart failure and shock with major CC
- 292: Heart failure and shock with CC
- 293: Heart failure and shock without CC/major CC
- 313: Chest pain
- 391: Esophagitis, gastroenteritis, and miscellaneous digestive disorders with major CC
- 392: Esophagitis, gastroenteritis, and miscellaneous digestive disorders without major CC
- 223: Cardiac defibrillator implant with cardiac catheterization with AMI/HF/shock without major CC
- 225: Cardiac defibrillator implant with cardiac catheterization without AMI/HF/shock without major CC
- 226: Cardiac defibrillator implant without cardiac catheterization with major CC

- 227: Cardiac defibrillator implant without cardiac catheterization without major CC
- 242: Permanent cardiac pacemaker implant with major CC
- 243: Permanent cardiac pacemaker implant with CC
- 244: Permanent cardiac pacemaker implant without CC/major CC
- 245: AICD (automatic implantable cardioverter defibrillator) generator procedures (246: Percutaneous cardiovascular procedure with drug-eluting stent with major CC or 4+ vessels/stents
- 248: Percutaneous cardiovascular procedure with non–drug-eluting stent with major CC or 4+ vessels/stents
- 249: Percutaneous cardiovascular procedure with non–drug-eluting stent without major CC
- 251: Percutaneous cardiovascular procedure without coronary artery stent or AMI without major CC
- 259: Cardiac pacemaker device replacement without major CC
- 261: Cardiac pacemaker revision except for device replacement with CC
- 262: Cardiac pacemaker revision except for device replacement without CC/major CC

## Scoring and Ranking

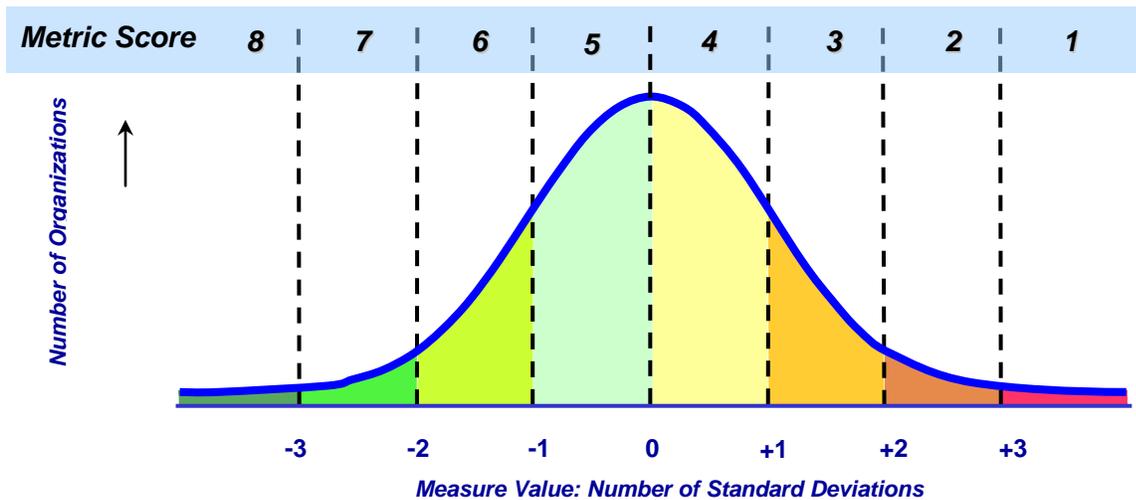
### *Score Calculations for Mortality, Safety, Efficiency, and Patient Centeredness*

Raw scores/rates for the metrics within each domain were transformed using an iterative approach to achieve a normal distribution. A test of normalcy was performed on each metric using the Shapiro-Wilk test. Each metric was tested for normalcy after each of the following iterative data manipulations:

- Exclusion of low volumes (described below in the section titled Missing Data/Low Volume)
- Exclusion of outliers ( $\pm 3$  standard deviations)
- Transformation using natural logarithm, square root, or cube root

Every metric could be normalized using this method. Normalization led to good dispersion of the raw rates. Once a measure was normally distributed, the mean and standard deviation were calculated for hospital-level performance. A score between 1 and 8 was assigned for each metric, with a point awarded for each standard deviation of movement from the mean in a favorable direction (Figure 1).

**Figure 1. Scoring Chart for a Measure for Which a Low Value Is Favorable**



***Score Calculation for Effectiveness***

The measures used to evaluate Effectiveness were derived from the Joint Commission Core Measures and all-cause 30-day readmission rate from the UHC CDB. As in previous years’ rankings, the 2011 core measures scores are based on actual performance rather than normalized data. The continuing improvement in core measures performance levels has led to clustering at high levels of compliance. Since the target for these measures is 100% compliance, the core measures bundles for AMI, HF, PN, and SCIP were scored on a fixed scale for actual bundle performance instead of a normalized distribution: 8 points for  $\geq 95\%$ , 7 points for  $\geq 90\%$ , 6 points for  $\geq 85\%$ , 5 points for  $\geq 80\%$ , 4 points for  $\geq 75\%$ , 3 points for  $\geq 70\%$ , 2 points for  $\geq 65\%$ , and 1 point for  $< 65\%$ . The 30-day all-cause readmission rate was normalized and then scored on the 8-point scale.

***Score Calculation for Equity***

The scoring method for Equity has been used since 2008. The purpose of the metric is to detect inequities in care to provide clear direction for efforts to resolve the inequities. The Fisher exact test was used to identify statistically significant differences in compliance rates among members for each dimension of equity (gender, race, and socioeconomic status). A conservative alpha level of 0.01 was used to define significant differences in core measure compliance by group to reduce Type I errors in the detection of an inequity. For each equity dimension, 4 core measure sets were evaluated and 4 Fisher exact tests were performed to determine whether an inequity existed between the groups in that dimension. Each of the 3 equity dimensions was worth a total of 8 points, for a possible score of 24. Each Fisher exact test on

each core measures set contributed 2 points—i.e., each Fisher exact test result with a  $P$  value  $< 0.01$  resulted in a deduction of 2 points from the maximum of 24 points.

### ***Missing Data/Low Volume***

Instead of imputing data to the median for metrics in the Mortality and Safety domains for individual hospitals with volumes too low to be scored reliably, UHC identified a volume threshold for each metric. For Mortality, the volume threshold for each product line was established individually based on the following calculation:  $1/(\text{standard deviation of the hospital-level observed mortality rate})$ . This calculated threshold is the volume of cases at which 1 death or safety-related event would move a hospital 1 standard deviation from the mean. Since the scoring for each is based on standard deviations, volumes less than these thresholds were considered too low to score reliably. For the Safety domain, the data were evaluated and a volume threshold requiring  $\geq 25$  cases in the denominator was determined. The domain score was calculated using only the scores for metrics that meet the volume thresholds and for which data were available.

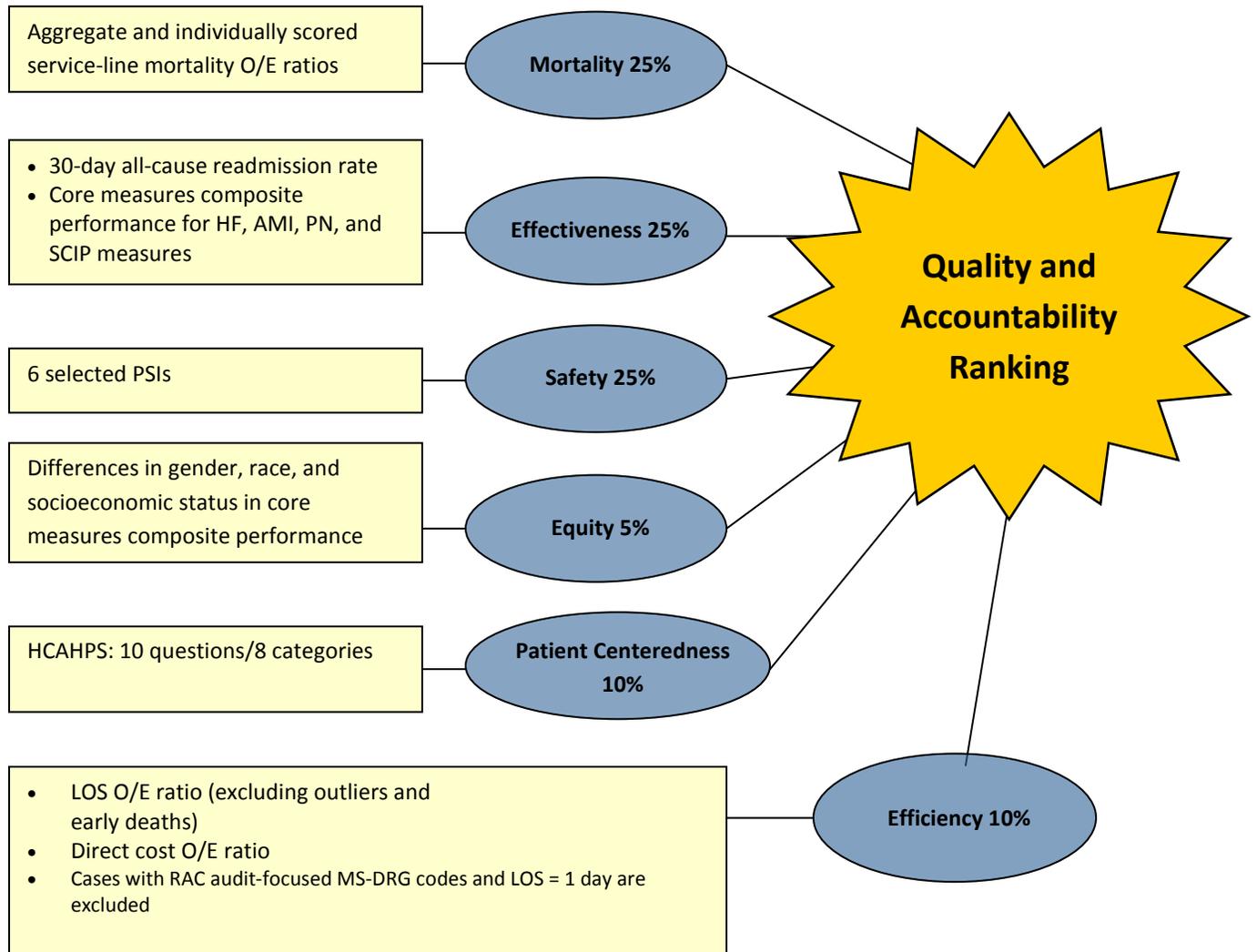
Both the Effectiveness and Equity measures relied on available core measures data. Twenty-six UHC members that do not submit core measures data to UHC were invited to submit data for inclusion in the scoring and ranking, and 23 did so. For those that did not submit their data and in cases in which an organization did not have data for a specific measure set, the median bundle rate for the entire cohort was assigned; 8 for AMI, 7 for HF, 6 for PN, and 6 for SCIP. A rate of 100% (no disparities at  $P < .01$ ) was imputed for the Equity domain, since the majority of hospitals had no disparities.. To take advantage of the risk and mode adjustment used by CMS in reporting HCAHPS results, data from the Hospital Compare Web site (Q4 2009 through Q3 2010) will be used for the Patient Centeredness domain. For any organizations that did not have data in the Hospital Compare database, the median score of 4.5 was assigned.

### ***Summarizing Scores***

The points awarded for each metric within the individual domains were added and the resulting total was divided by the maximum total points possible for those domains in which the institution had sufficient data. That percentage is the hospital's score for the individual domain. Ranks for each domain were created based on the hospital's percentage score; institutions could be tied in rank for a particular domain. The Mortality, Effectiveness, and Safety domains are weighted equally at 25% each. In previous years, to adjust for any potential PSI definitional issues or differences in coding practices, the Safety domain was weighted less. However, UHC believes that the use of the PSIs nationally has continued to grow and that the availability of the present-on-admission flags improves the validity and accuracy of these metrics. The steering committee recognizes that while the PSIs are still not perfect indicators of safety performance, they are currently the best safety metrics that are widely available. Thus, in 2011, the Safety domain is

weighted the same as Mortality and Effectiveness. The Patient Centeredness and Efficiency domains are weighted at 10% each, and Equity is weighted at 5%. Figure 2 shows the ranking methodology used, including the relative weights given to the domains.

**Figure 2. Domain Weighting Used for Organizational Score**



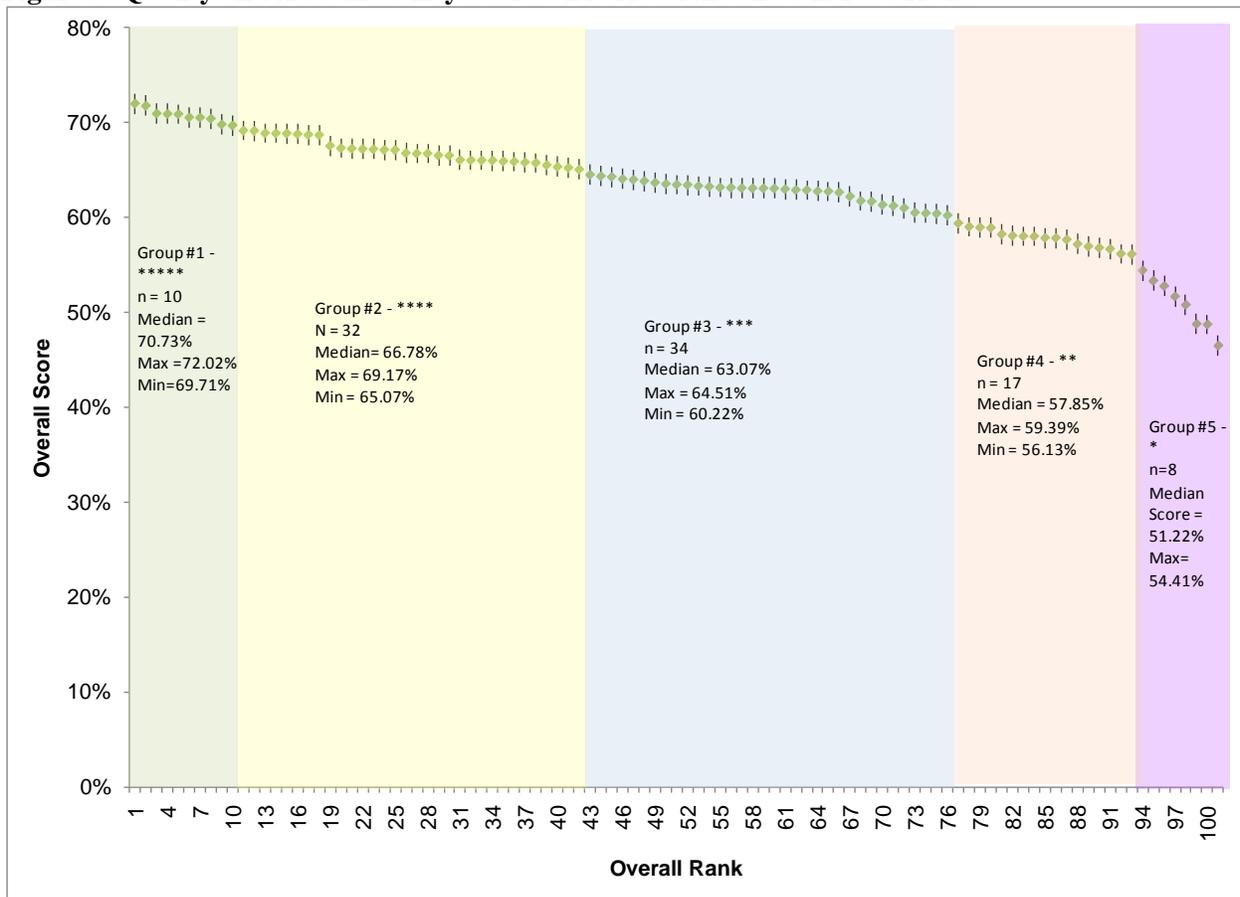
AMI = acute myocardial infarction; HCAHPS = Hospital Consumer Assessment of Healthcare Providers and Systems; HF = heart failure; LOS = length of stay; O/E = observed to expected; PN = pneumonia; PSI = Agency for Healthcare Research and Quality Patient Safety Indicator; RAC = Recovery Audit Contractor; SCIP = Surgical Care Improvement Project.

The weighted scores for each domain were added to arrive at the overall score, and then each institution was assigned a rank based on this overall score. Because the final score is a composite, it is possible for an institution to have a high overall score but still show considerable opportunity for improvement in a

particular domain. In addition, because the overall score is a comparative ranking, organizations may see significant organizational improvement but may not see the same incremental improvement in the ranking as other organizations that also improve their performance.

Organizations are grouped by score into 5 categories based on the minimum overlap in 95% confidence interval bands around each score—for 2010, there were 10 organizations in group 1, 32 in group 2, 34 in group 3, 17 in group 4, and 8 in group 5. A summary of the data for each group is shown in Figure 3.

**Figure 3. Quality and Accountability Scores: Distributions and Characteristics**



## Discussion

The approach of normalizing the data continues to make it extremely difficult for any institution to obtain a perfect score. To achieve 100%, performance would need to be at least 3 standard deviations from the mean in a favorable direction for each Safety, Mortality, Efficiency, and Patient Centeredness metric. In addition, core measure composite compliance would need to be in excess of 95% on each measure set, and there could be no statistically significant disparities in the Equity domain. A subset of the full list of PSIs was used for the scoring. AHRQ and other organizations, including UHC, continue to assess the

reliability and sensitivity of the PSIs, and others are expected to be included in future iterations of the ranking as specificity and sensitivity improve.

UHC recognizes that the methodology for the Quality and Accountability Study changes each year. The steering committee debated the extent of these changes, finally agreeing that a careful balance should be maintained between improving the methodology and keeping it consistent over time to allow for performance monitoring. This decision was made to better capture true differences in quality and safety performance and to reflect the broadening scope of nationally reported safety measures. Each year when the methodology for the study is finalized, UHC applies the revised methodology to previous periods to evaluate changes in performance over time and shares this information with interested organizations.

### **Limitations of the Model**

This model is designed to compare organizations' performance on selected measures, with scoring distinctions based on standard deviations from a given mean. Therefore an organization may have made notable performance improvements from one year to the next and yet still remain within a standard deviation band and achieve the same overall score. This is particularly true for institutions scoring in the middle range ( $\pm 1$  SD) on a metric.

The ability to accurately gauge Effectiveness and Equity performance depends on the availability of patient-level core measures data. The rankings of any organizations that do not submit their data to UHC may not accurately reflect their true position among UHC members since data are imputed in those cases.

The population base for the study is limited to full UHC members that are active in UHC's Clinical Data Base (including several members new to UHC and/or the CDB). The optimal combination of measures, domains, and weights may be different for nonacademic community medical centers as well as for affiliate hospitals associated with UHC member institutions; therefore, this methodology may not be applicable to nonmembers or associate members.

AHRQ's PSIs were developed with a view to face validity (agreement that they represent an important component of quality and safety) and construct validity (consistency with other available measures of quality and safety). Many have not been tested extensively for sensitivity and specificity, though research efforts in this area continue to progress. False positives may occur because of failure to distinguish between complications of care and comorbidities present at hospital admission, though the widespread use of "present on admission" flags in discharge abstracts has made such false positives less common. False negatives may occur because of failure to capture or document the complications in question. Rates

may vary because of differences in coding practices or case mix. The newly issued risk-adjustment software from AHRQ is intended to address the latter issue. Despite these caveats, the AHRQ PSIs remain the best measures of safety currently available from administrative data sets.

UHC will continue to examine and refine its methods, based on feedback from member organizations, the emergence of new performance measures and domains, and the maturation of long-standing measures that lead to performance coalescing in a narrow range.

### **Acknowledgements**

This year UHC requested an expert review of the statistical approaches used to transform the data and the SAS programs that are the foundation for the ranking. Special thanks to Vinita Bahl, DMD, MPP, director, Clinical Information & Decision Support Services, and Hsou Mei (May) Hu, PhD, MBA, MHS, senior clinical information analyst, at University of Michigan Hospitals & Health Centers, for their review, feedback, and enhancements to ensure a sound methodology.

## **Appendix A. Committee Members**

### **Quality and Accountability Study Steering Committee**

- Vinita Bahl, DMD, MPP, director, Clinical Information & Decision Support Services, University of Michigan Hospitals & Health Centers
- Julie Cerese, RN, MSN, vice president, Performance Improvement, UHC
- Maureen Disbot, MS, RN, CCRN, vice president, Quality Operations, The Methodist Hospital
- John Duval, chief executive officer, MCV Hospitals
- David Entwistle, chief executive officer, University of Utah Hospitals & Clinics
- Jonathan Gottlieb, senior vice president and chief medical officer, University of Maryland Medical Center
- Raj Iyer, manager, Product Development, UHC
- Mark Keroack, MD, MPH, chief physician executive, Baystate Health, Inc (formerly senior vice president and chief medical officer, UHC)
- Lee Norman, MD, MHS, MBA, senior vice president and chief medical officer, The University of Kansas Hospital Authority
- Leslie Prellwitz, MBA, director, Performance Improvement Analytics, UHC
- Martha Radford, MD, FACC, FAHA, chief quality officer, NYU Langone Medical Center
- Allison Sabel, MD, PhD, MPH, director of biostatistics and clinical data warehouse, Denver Health
- Danny Sama, MBA, manager, Clinical Quality & Analytics, Northwestern Memorial Hospital
- Carolyn Sanders, vice president for patient services and chief nursing officer, University of Colorado Hospital
- Kevin Tabb, MD, chief medical officer, Stanford Hospital & Clinics
- Brian Taylor, PhD, director, Clinical Analytics, NewYork-Presbyterian Hospital

### **Performance Improvement/Comparative Data & Informatics Committee of the UHC Board of Directors**

- Tim Babineau, MD, president and chief executive officer of Rhode Island Hospital and Miriam Hospital
- Bernard Birnbaum, MD, senior vice president, vice dean, and chief of hospital operations, NYU Langone Medical Center
- Marc Boom, MD, executive vice president, The Methodist Hospital
- Lynn Frederick, hospital administrator, Mayo Clinic—Saint Mary's Hospital

- Richard Lofgren, MD, vice president, Healthcare Operations, chief clinical officer, University of Kentucky Hospital
- Tom McAfee, dean of clinical affairs, UC San Diego Health System
- Sam Odle, executive vice president/chief operating officer, Indiana University Health
- Bob Page, president and chief executive officer, The University of Kansas Hospital Authority
- Gary Park, president, University of North Carolina Hospitals
- Amir Rubin, president and chief executive officer, Stanford Hospital & Clinics
- Bruce Schroffel, president and chief executive officer, University of Colorado Hospital
- Kevin Sowers, chief executive officer, Duke University Hospital
- Johnese Spisso, chief health system officer, UW Medicine Health System

## Appendix B: 2011 Quality and Accountability Study Metrics, by Domain

### Domain 1: Mortality

<i>Source: UHC Clinical Data Base</i>			
<i>Time frame: Discharges from July 2010 through June 2011</i>			
<i>Overall exclusions: Patients with an admit source of transfer from hospice, nonviable neonates, cases flagged as bad data</i>			
<b>Metric</b>	<b>Metric Used</b>	<b>Transformations Used</b>	<b>Scoring</b>
<b>Individually Reviewed Service Lines</b>			
Cardiology	Mortality O/E	Outlier	Each service line is scored and ranked individually; composite counts for half of Mortality domain score
Gastroenterology	Mortality O/E	Outlier	
Medical oncology	Mortality O/E	Outlier	
General medicine	Mortality O/E	None	
Neurology	Mortality O/E	Outlier	
General surgery	Mortality O/E	Outlier	
Neurosurgery	Mortality O/E	Square root	
Cardiothoracic surgery (cardiac surgery and thoracic surgery service lines)	Mortality O/E	Square root	
<b>Aggregate Service Lines</b>			
Bone marrow transplant Burns Cardiology Cardiothoracic surgery Gastroenterology Gynecology Gynecologic oncology Heart/lung transplant <sup>a</sup> HIV Kidney/pancreas transplant Liver transplant Medical oncology Medicine, general Neurology Neurosurgery Obstetrics Orthopedics Otolaryngology Plastic surgery Rheumatology Spinal surgery Surgical oncology Surgery, general Trauma Urology Vascular surgery Ventilator support	Mortality O/E	Outlier	Mortality O/E is calculated for these lines as a group; composite counts for half of Mortality domain score

<sup>a</sup> Only discharges with ICD-9 procedure codes 336 or 3751 are included; excludes cases that only involve a heart-assist system. ICD-9 = *International Classification of Diseases, 9th Revision*; O/E = observed-to-expected ratio.

## Domain 2: Safety

Source: UHC Clinical Data Base

Time Frame: Discharges from July 2010 through June 2011

Overall exclusions: Patients with an admit source of transfer from hospice, nonviable neonates, cases flagged as bad data

Metric	Metric Used	Transformations Used	Scoring
PSI-3: Pressure ulcer (AHRQ v. 3.2)	O/E	Square root	Measures individually scored; each counts equally in overall Safety domain score
PSI-6: Iatrogenic pneumothorax (AHRQ v. 4.2)	O/E	Outlier	
PSI-7: Central line-associated bloodstream infection (AHRQ v. 4.2)	O/E	Square root	
PSI-9: Postoperative hemorrhage and hematoma (AHRQ v. 4.2)	O/E	Log	
PSI-11: Postoperative respiratory failure (AHRQ v. 4.2)	O/E	Log	
PSI-12: Postoperative pulmonary embolism or deep vein thrombosis (AHRQ v. 4.2)	O/E	Log	

AHRQ = Agency for Healthcare Research and Quality; O/E = observed-to-expected ratio; PSI = AHRQ Patient Safety Indicator.

### Domain 3: Effectiveness

Source: UHC Core Measures database

Time frame: Discharges from April 2010 through March 2011

Metric	Metrics Used	Transformations Used	Scoring
Core measures—AMI composite	AMI-1, AMI-2, AMI-3, AMI-4, AMI-5, AMI-7a, AMI-8a	N/A	Absolute performance scale used for each composite based on compliance:  8 points for $\geq 95\%$ 7 points for $\geq 90\%$ 6 points for $\geq 85\%$ 5 points for $\geq 80\%$ 4 points for $\geq 75\%$ 3 points for $\geq 70\%$ 2 points for $\geq 65\%$ 1 point for $< 65\%$
Core measures—PN composite	PN-2, PN-3a, PN-3b, PN-4, PN-5c, PN-6, PN-7	N/A	
Core measures—HF composite	HF-1, HF-2, HF-3, HF-4	N/A	
Core measures—SCIP composite	SCIP-Inf-1a, SCIP-Inf-2a, SCIP-Inf-3a, SCIP-Inf-4, SCIP-Inf-6, SCIP-Inf-9, SCIP-Inf-10, SCIP-Card-2, SCIP-VTE-1, SCIP-VTE-2	N/A	
Readmissions	30-day all-cause  Exclusions: From denominator: obstetrics, neonatology, and normal newborns service lines; death at 1st admission  From numerator: dialysis, delivery, rehabilitation, radiation therapy, chemotherapy, psychiatric (within 1 day)  Global exclusions: Patients with an admit source of transfer from hospice, nonviable neonates, cases flagged as bad data	None	1-8 based on standard deviations

AMI = acute myocardial infarction; Card = cardiac; HF = heart failure; Inf = infection; N/A = not applicable; PN = pneumonia; SCIP = Surgical Care Improvement Project; VTE = venous thromboembolism.

## Domain 4: Patient Centeredness

Source: Hospital Compare database

Time frame: Discharges between October 2009 and September 2010

Metric	Transformations Used?	Scoring
<b>Category:</b> Nurse Communication <ul style="list-style-type: none"> <li>• How often did nurses communicate well with patients?</li> </ul>	Outlier	Each category was scored and ranked individually on either the percentage responding with an affirmative (“yes” for yes/no questions), or superlative (“always”) response
<b>Category:</b> Doctor Communication <ul style="list-style-type: none"> <li>• How often did doctors communicate well with patients?</li> </ul>	None	
<b>Category:</b> Pain Management <ul style="list-style-type: none"> <li>• How often was patients’ pain well controlled?</li> </ul>	None	
<b>Category:</b> Communication About Medications <ul style="list-style-type: none"> <li>• How often did staff explain about medicines before giving them to patients?</li> </ul>	None	
<b>Category:</b> Cleanliness and Quietness <ul style="list-style-type: none"> <li>• How often was the area around patients’ rooms kept quiet at night?</li> <li>• How often were the patients’ rooms and bathrooms kept clean?</li> </ul>	None	
<b>Category:</b> Responsiveness of Hospital Staff <ul style="list-style-type: none"> <li>• How often did patients receive help quickly from hospital staff?</li> </ul>	None	
<b>Category:</b> Discharge Information <ul style="list-style-type: none"> <li>• Were patients given information about what to do during their recovery at home?</li> </ul>	Cube root	
<b>Category:</b> Overall Rating of Hospital <ul style="list-style-type: none"> <li>• How do patients rate the hospital overall?</li> <li>• Would patients recommend the hospital to friends and family?</li> </ul>	None	

## Domain 5: Equity

Source: UHC Core Measures database combined with UHC Clinical Data Base

Time frame: Discharges between April 2010 and March 2011

Differences in gender (male/female), race (white/nonwhite), and socioeconomic status (Medicaid, county medically indigent, charity, self-pay/uninsured, and Title V maternal/child health/all other payer types) in core measures composite performance

Metric	Inclusions/Exclusions	Transformations Used	Scoring
AMI composite—gender	None other than those for core measures effectiveness metrics	N/A	Fisher's exact test; each test with result $P < 0.01$ deducts 2 points from maximum of 24
AMI composite—race			
AMI composite—socioeconomic status			
HF composite—gender			
HF composite—race			
HF composite—socioeconomic status			
SCIP composite—gender			
SCIP composite—race			
SCIP composite—socioeconomic status			
PN composite—gender			
PN composite—race			
PN composite—socioeconomic status			

AMI = acute myocardial infarction; HF = heart failure; N/A = not applicable; PN = pneumonia; SCIP = Surgical Care Improvement Project.

## Domain 6: Efficiency

Source: UHC Clinical Data Base

Time period: Discharges between July 2010 and June 2011

Overall Exclusions: Patients with an admit source of transfer from hospice, nonviable neonates, cases flagged as bad data

Metric	Inclusions/Exclusions	Transformations Used	Scoring
LOS O/E	<p>Inclusions: Cases in these service lines:</p> <ul style="list-style-type: none"> <li>• Cardiology</li> <li>• Gastroenterology</li> <li>• Medical oncology</li> <li>• General medicine</li> <li>• Neurology</li> <li>• General surgery</li> <li>• Neurosurgery</li> <li>• Cardiac and thoracic surgery</li> </ul> <p>Exclusions: Cases within the above service lines with LOS = 1 day and MS-DRG codes 069, 190, 191, 192, 291, 292, 293, 313, 391, 392, 223, 225, 226, 227, 242, 243, 244, 245, 246, 248, 249, 251, 259, 261, 262</p>	Outlier	LOS O/E calculated as one large group, similarly for Direct Cost O/E. Each composite counts equally in the overall domain score.
Direct cost O/E		Log	

LOS = length of stay; O/E = observed-to-expected ratio.