

Above & Beyond: the Power of
Minds in Motion

UHC 2010 QUALITY & SAFETY FALL FORUM

UHC 2010 Quality and Accountability Study

Scoring and Ranking Methodology

September 2010

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.

In 2010, UHC expanded upon the original 2005 study by researching characteristics associated with organizations that have transformed into high performers. The preliminary results of this new study were presented at the UHC 2010 Quality & Safety Fall Forum.

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 ranking 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 and results of the 2010 Quality and Accountability rankings.

Methods

Principles of Structural Development

As in 2005, the 2010 rankings were 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. Metrics for efficiency are also provided on organizations' scorecards to give an additional perspective on organizational success; however, these were not factored into the overall institutional rankings.

Timeliness data, focusing on organizations' 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; 98 member institutions were included in the 2010 analysis. Source data were obtained from UHC's Clinical Data Base (CDB), Operational Data Base (ODB), 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 Clinical Evaluative Sciences Council steering committee and member experts in analysis of performance data. (See Appendix A for the members of the Quality and Accountability Study Steering Committee.)

Data Sources, Domains, and Measures

Institutional performance metrics were grouped into 6 domains: Mortality, Effectiveness, Safety, Equity, Patient Centeredness, and Efficiency. The first 5 domains were used to calculate each hospital's overall score and rank; efficiency data were scored and reported on the scorecard but not included in the final score calculations. To evaluate sustained performance and to 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 2009 through Q2 2010 for CDB and Q2 2009 through Q1 2010 for Core Measures and ODB) were used. Patients with an admit source of transfer from hospice and nonviable neonates were excluded from the mortality, safety, readmission, and length of stay (LOS) metrics. The most recent HCAHPS data available

from the Hospital Compare Web site (Q4 2008 through Q3 2009) were used for the Patient Centeredness domain.

Mortality

The 2010 Mortality domain was scored using a hybrid approach to account for both system-level and product line-level performance. The rationale for this approach is that factors affecting the risk of death occur at both the product-line level and the system level. The Mortality domain score has 2 equal components: an aggregate mortality score and a composite of individually scored product lines. The aggregate mortality score is the observed-to-expected (O/E) mortality ratio for all patients in 28 UHC product lines; however, because of the overlap of diagnosis-related group (DRG) codes, 2 of the 27, “heart transplant or implant of heart assist system” and “lung transplant,” were combined into a single product line, “heart/lung transplant.” The final 27 product lines scored for the aggregate mortality component were:

- 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 product lines were reviewed and evaluated. Several very large product lines with significant mortality continue to be the best indicators of performance across the membership. The 7 product lines selected are those considered to be both large and strategic services for

an academic medical center and therefore provide a good opportunity to evaluate performance across members at the product-line level:

- Cardiology
- Cardiothoracic surgery
- Gastroenterology
- Medical oncology
- Medicine, general
- Neurology
- Surgery, general

Observed-to-expected mortality ratios for each of these 7 product lines were determined and weighted equally in calculating the product-line score. The product-line score and the aggregate score contributed equally to the final mortality score. All mortality calculations used UHC's risk-adjustment models, which are based on Medicare severity DRGs (MS-DRGs).

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 2010 rankings, because of the intense national focus on readmissions as defined by 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 are excluded from the numerator of the metric.

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 the Centers for Medicare & Medicaid Services (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-7, 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 9 that showed sufficient variation and incidence were selected for inclusion in the 2010 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
- PSI-17: Birth trauma—injury to neonate
- PSI-18: Obstetric trauma—vaginal delivery with instrument
- PSI-19: Obstetric trauma—vaginal delivery without instrument

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 a high signal ratio are still subject to high false-positive rates; therefore, all PSIs are evaluated annually for potential inclusion in this domain. In 2010, two PSIs—postoperative hemorrhage and hematoma and postoperative sepsis—were reviewed as candidates for inclusion, but because of the variability of the data within these measures, the steering committee recommended that they not be included this year. PSI-15: Accidental puncture or laceration, which had been used in previous rankings, was excluded this year at the recommendation of the steering committee because of the variability in interpretation; accidental puncture or laceration is often coded for intentional puncture. The 3 obstetric PSIs are scored at one-third the weight of the other PSIs.

Risk-adjusted rates based on AHRQ's PSI version 4.1 were used for PSI-6: iatrogenic pneumothorax; PSI-7: central line associated bloodstream infections; PSI-9: postoperative hemorrhage and hematoma; PSI-11: postoperative respiratory failure; and PSI-12: postoperative pulmonary embolism or deep vein thrombosis. For PSI-17: birth trauma – injury to neonate; PSI-18: obstetric trauma—vaginal delivery with instrument; and PSI-19: obstetric trauma—vaginal delivery without instrument, the AHRQ version 4.1

observed rate was used. For PSI-3: pressure ulcer, the AHRQ PSI version 3.2 risk-adjusted rate was used because version 4.1 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

Historically, 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?"—served as the source for the Patient Centeredness metric. In 2010, 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). The most recent 12 months of data available from this site were for the period Q4 2008 through Q3 2009.

Efficiency

Domain scores for Efficiency are shown on the Quality and Accountability Scorecard; however, these metrics were not used in calculating the rankings because they are not consistently available for a high percentage of UHC members. Efficiency metrics focused on per-case performance on cost and LOS. Cost data are from the UHC ODB and LOS data are from the UHC CDB. Measures include:

- Total expense per case mix index (CMI)–adjusted discharge net bad debt (wage index adjusted)
- Labor expense per CMI-adjusted discharge (wage index adjusted)
- Supply expense per CMI-adjusted discharge
- LOS observed/expected ratio (excluding LOS outliers and early deaths)

Scoring and Ranking

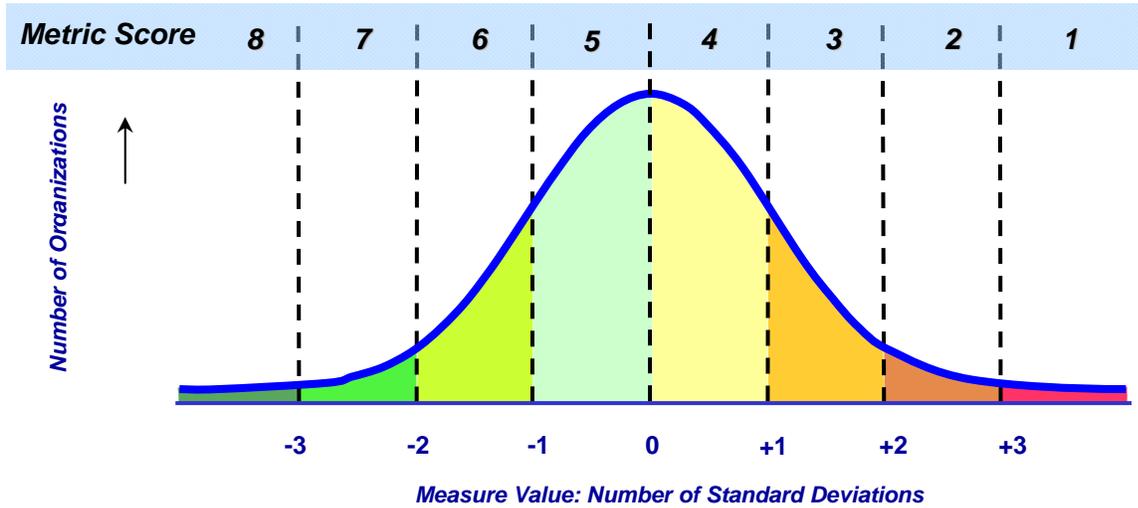
Score Calculations for Mortality, Safety, Efficiency, and Patient Centeredness

The scores for Mortality, Safety, Patient Centeredness, and Efficiency were determined as follows: 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 (± 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 the 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 rankings, the 2010 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 > 90%, 7 points for > 80%, 6 points for > 70%, 5 points for > 60%, 4 points for > 50%, 3 points for > 40%, 2 points for > 30%, and 1 point for ≤ 30%. The 30-day related readmission rate was normalized and then scored on the 8-point scale as was done with Mortality and Safety.

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. As was observed in 2008, the statistical testing for gender within the SCIP core measure revealed an unusually large number of disparities. In general, compliance with SCIP measures tends to be higher in the gynecology product line, which skews the performance scores for hospitals with either large or small volumes of gynecology cases. To eliminate skewing of results due to the effect of this gender-specific product line, the gynecology product line was removed from the analysis of gender equity in SCIP.

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 Safety, the data were evaluated and a volume threshold requiring ≥ 25 cases in the denominator was determined. The domain score was calculated using only the scores for metrics that met the volume thresholds and for which data were available.

Both the Effectiveness and Equity measures relied on available core measures data. Twenty-eight UHC members that do not submit core measures data to UHC were invited to submit data for inclusion in the scoring and ranking, and 26 did so. For the 2 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, 7 for PN, and 7 for SCIP. A rate of 100% (no disparities at $P < 0.01$) was imputed for the Equity domain, since the majority of hospitals had no disparities.

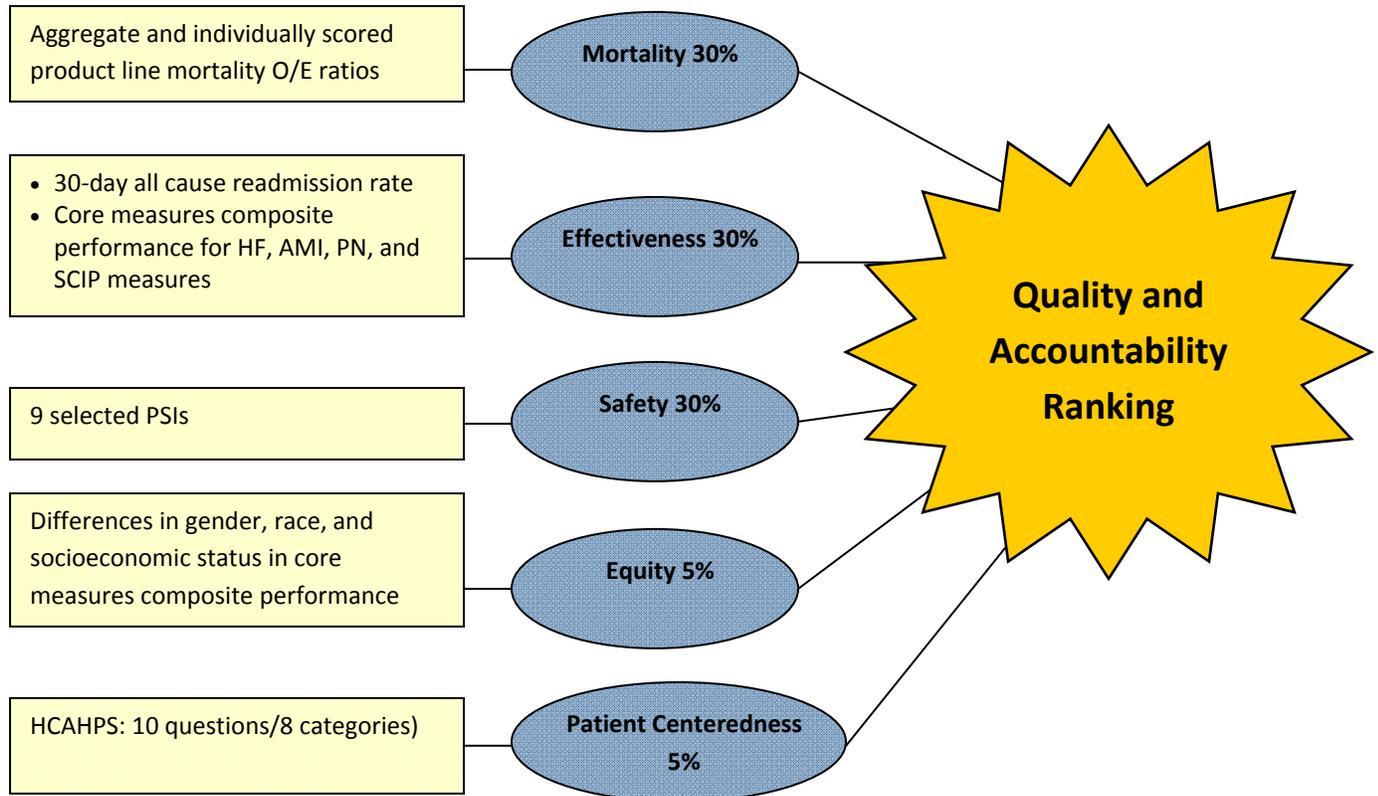
To take advantage of the risk and mode adjustment used by CMS in reporting HCAHPS results, data from the Hospital Compare Web site (Q4 2008 through Q3 2009) were used for the Patient Centeredness domain. Ninety-six institutions in the 2010 ranking submitted HCAHPS data for public reporting on Hospital Compare. For the 2 organizations that did not have data in the Hospital Compare database, the median score of 5 was assigned.

The UHC ODB was the data source for all but the LOS metric in the Efficiency domain. For hospitals that do not participate in the ODB, the median score was imputed (total cost, 5; labor cost, 4; supply cost, 5).

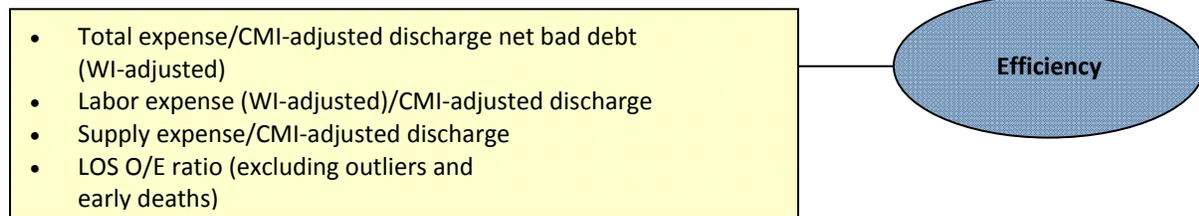
Summarizing Scores

The points awarded for each metric within the individual domains were added; the resulting total was divided by the maximum total points possible for those domains in which the institution had sufficient data. That percentage was 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 were weighted equally at 30% each. In previous years, to adjust for any potential PSI definitional issues or differences in coding practices, the Safety domain was weighted less. UHC feels that the use of the PSIs nationally continues to grow and that the availability of the present-on-admission flags improves the validity and accuracy of these metrics. The steering committee continues to feel that the PSIs are still not perfect indicators of safety performance, but agrees that they are currently the best safety metrics that are widely available. Thus the Safety domain was weighted the same as Mortality and Effectiveness. The Equity and Patient Centeredness domains were weighted at 5% each. Figure 2 shows the ranking methodology used, including the relative weights given to the domains.

Figure 2. Domain Weighting Used for Organizational Score



Additional Measures of Organizational Success

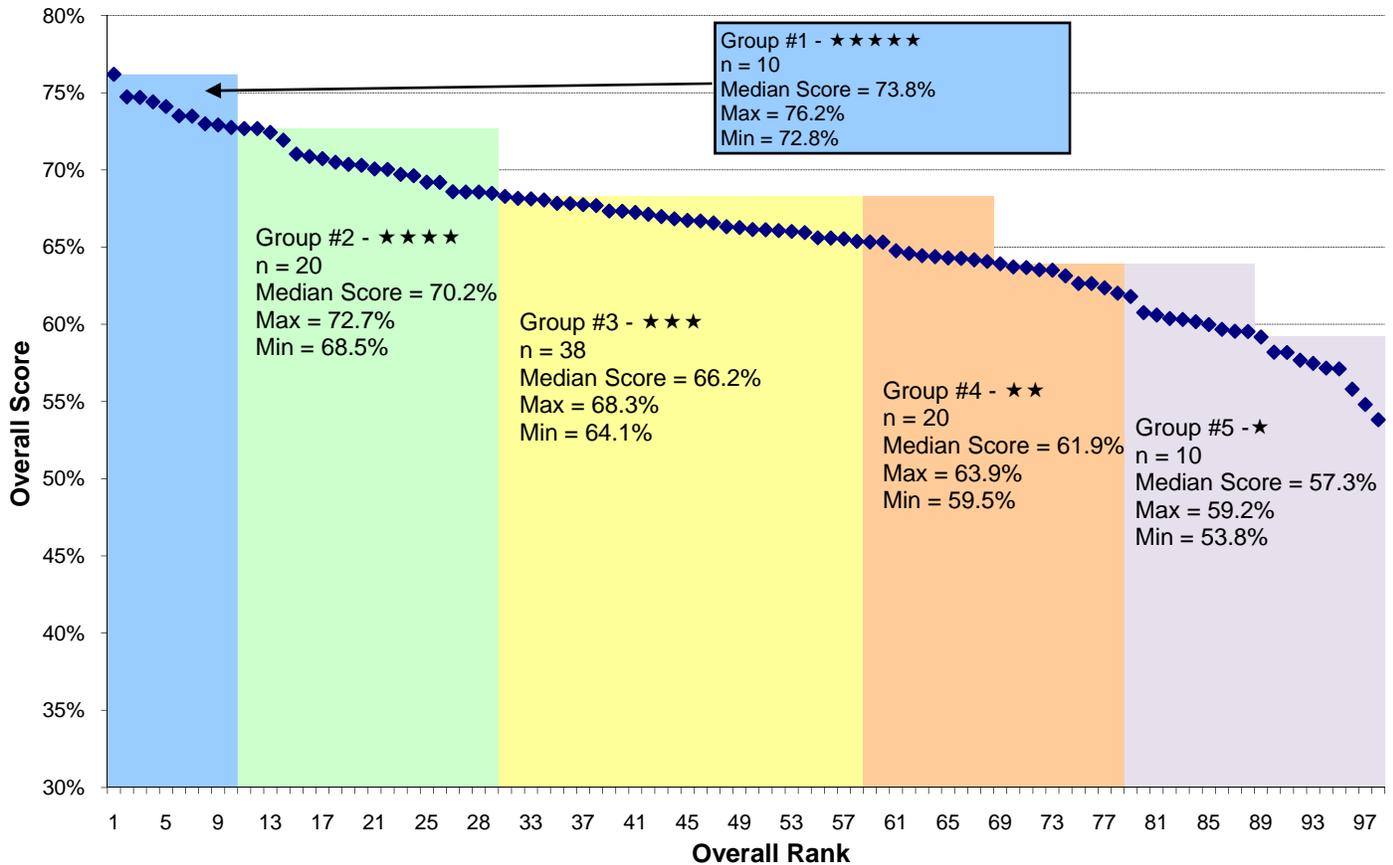


AMI = acute myocardial infarction; CMI = case mix index; HCAHPS = Hospital Consumer Assessment of Healthcare Providers and Systems; HF = heart failure; LOS = length of stay; O/E = observed to expected; PN = pneumonia; PSI = Agency for Healthcare Research and Quality Patient Safety Indicator; SCIP = Surgical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

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 also improve their performance. Organizations were grouped by score into 5 arbitrary categories—10 organizations in group 1, 20 in group 2, 38 in group 3, 20 in group 4, and 10 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 and Mortality metric. Although a perfect score is difficult to achieve, the median score for the top tier of hospitals increases each year because of improvements in core measures performance.

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 rankings changes each year. The steering committee debated the extent of these changes, finally agreeing that a careful balance should be maintained between keeping the methodology consistent over time to allow for performance monitoring and improving the methodology to better capture true differences in quality and safety performance and 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 (± 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 the 2 organizations that did not submit their data to UHC may not accurately reflect their true position among UHC members since data had to be imputed.

The population base for the study was 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 overall methodology, 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. Quality and Accountability Study Steering Committee

UHC would like to thank the members of the steering committee, who provided oversight of the development of the methodology for the 2010 Quality and Accountability rankings, for their invaluable time and insight.

- Vinita Bahl, DMD, MPP, director, Clinical Information & Decision Support Services, University of Michigan Hospitals & Health Centers
- John Brumsted, MD, chief medical officer, Fletcher Allen Health Care
- William Burton, senior director, Performance Services, Duke University Health System
- Julie Cerese, RN, MSN, vice president, Performance Improvement, University HealthSystem Consortium
- Bruce Davidson, PhD, MPH, director, Resource & Outcomes Management, Cedars-Sinai Medical Center
- Maureen Disbot, MS, RN, CCRN, vice president, Quality Operations, The Methodist Hospital
- Raj Iyer, manager, Product Development, University HealthSystem Consortium
- Victoria Jordan, PhD, MBA, MS, director, Quality Engineering and Clinical Operations Informatics, The University of Texas MD Anderson Cancer Center
- Mark Keroack, MD, MPH, senior vice president and chief medical officer, University HealthSystem Consortium
- Steve Meurer, PhD, MBA, MHS, senior vice president, Comparative Data & Informatics, University HealthSystem Consortium
- Lee Norman, MD, MHS, MBA, senior vice president and chief medical officer, The University of Kansas Hospital Authority
- 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
- Kevin Tabb, MD, chief medical officer, Stanford Hospital & Clinics
- Brian Taylor, PhD, director, Clinical Analytics, NewYork-Presbyterian Hospital