



Published in final edited form as:

Ann Thorac Surg. 2019 February ; 107(2): 590–596. doi:10.1016/j.athoracsur.2018.07.036.

Development of a Congenital Heart Surgery Composite Quality Metric: Part 2—Analytic Methods

Sean M. O'Brien, PhD¹, Jeffrey P. Jacobs, MD², David M. Shahian, MD³, Marshall L. Jacobs, MD², J. William Gaynor, MD⁴, Jennifer C. Romano, MD⁵, Michael G. Gaies, MD, MPH⁶, Kevin D. Hill, MD, MS¹, John E. Mayer, MD⁷, and Sara K. Pasquali, MD, MHS⁸

¹Duke Clinical Research Institute, Duke University School of Medicine, Durham, North Carolina.

²Division of Cardiac Surgery, Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland; Division of Cardiovascular Surgery, Department of Surgery, Johns Hopkins All Children's Heart Institute, St. Petersburg, Florida.

³Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts.

⁴Department of Surgery, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.

⁵Department of Cardiac Surgery, University of Michigan Medical School, Ann Arbor, Michigan.

⁶Department of Pediatrics, University of Michigan C.S. Mott Children's Hospital, Ann Arbor, Michigan.

⁷Department of Cardiovascular Surgery, Boston Children's Hospital, Boston, Massachusetts.

⁸Department of Pediatrics, University of Michigan C.S. Mott Children's Hospital, Ann Arbor, Michigan.

Abstract

Background—We describe the statistical methods and results related to development of the first congenital heart surgery composite quality measure.

Methods—The composite measure was developed using The Society of Thoracic Surgeons Congenital Heart Surgery Database (2012 to 2015), Bayesian hierarchical modeling, and the current Society of Thoracic Surgeons risk model for case-mix adjustment. It consists of a mortality domain (operative mortality) and morbidity domain (major complications and postoperative length of stay). We evaluated several potential weighting schemes and properties of the final composite measure, including reliability (signal-to-noise ratio) and hospital classification in various performance categories.

Address correspondence to Dr Pasquali, C.S. Mott Children's Hospital, 1540 E Hospital Dr, Ann Arbor, MI 48105; pasquali@med.umich.edu.

Presented at the Fifty-fourth Annual Meeting of The Society of Thoracic Surgeons, Fort Lauderdale, FL, Jan 27–31, 2018. Richard E. Clark Memorial Paper for Congenital Heart Surgery.

The Appendix can be viewed in the online version of this article [<https://doi.org/10.1016/j.athoracsur.2018.07.036>] on <http://www.annalsthoracicsurgery.org>.

Results—Overall, 100 hospitals (78,425 operations) were included. Each adjusted metric included in the composite varied across hospitals: operative mortality (median, 3.1%; 10th to 90th percentile, 2.1% to 4.4%) major complications (median 11.7%, 10th to 90th percentile, 6.4% to 17.4%), and length of stay (median, 7.0 days; 10th to 90th percentile, 5.9 to 8.2 days). In the final composite weighting scheme selected, mortality had the greatest influence, followed by major complications and length of stay (correlation with overall composite score of 0.87, 0.69, and 0.47, respectively). Reliability of the composite measure was 0.73 compared with 0.59 for mortality alone. The distribution of hospitals across composite measure performance categories (defined by whether the 95% credible interval overlapped The Society of Thoracic Surgeons average) was 75% (same as expected), 9% (worse than expected), and 16% (better than expected).

Conclusions—This congenital heart surgery composite measure incorporates aspects of both morbidity and mortality, has clinical face validity, and greater ability to discriminate hospital performance compared with mortality alone. Ongoing efforts will support the use of the composite measure in benchmarking and quality improvement activities.

Quality measures are used across multiple settings in pediatric and congenital heart surgery, including benchmarking and quality improvement activities, public reporting, and designation of centers of excellence by certain payers. Current quality measures focus primarily on operative mortality, which has several limitations. As described in detail in Part 1 of this report, these include its unidimensional nature and limited ability to discriminate hospital performance due to low overall event rates in the current era [1].

More comprehensive measures incorporating additional components of quality are important to numerous stakeholders and hold the potential to augment the variety of improvement, benchmarking, and reporting efforts described above. Part 1 of this report detailed the background, rationale, and conceptual framework related to the development of the first composite quality metric in pediatric and congenital heart surgery [1]. In this report (Part 2), we describe the statistical methods and results.

Patients and Methods

This study was approved by the Duke University and University of Michigan Institutional Review Boards and was not considered human subjects research in accordance with the Common Rule (45 CFR 46.102(f)).

Data Source

The Society of Thoracic Surgeons Congenital Heart Surgery Database (STS-CHSD) collects standardized perioperative data on all patients undergoing pediatric and congenital heart operations at participating hospitals [2, 3]. Data quality is optimized through data checks, site visits, and audits [3].

Study Population

Patients undergoing any index cardiovascular operation, with or without cardiopulmonary bypass, at North American hospitals participating in the STS-CHSD from 2012 to 2015 were included (117 hospitals and 100,203 operations). Only the first (index) cardiovascular

operation of each hospital admission was analyzed. We excluded 16 hospitals that had more than 10% missing data for key variables and 1 hospital with an extremely low incidence of complications, presumed to be inaccurate. From the remaining 100 hospitals (86,154 operations) we excluded 4,018 infants weighing less than 2.5 kg undergoing isolated ductus arteriosus ligation, 1,806 operations without an STS-European Association for Cardiothoracic Surgery (STAT) mortality score, 767 records with data collected under an obsolete data collection form, 851 records with missing data for operative mortality or complications, and 287 records with missing data for age, sex, or weight. The final cohort consisted of 78,425 operations (100 hospitals).

We further examined the influence of heart and lung transplant procedures (Appendix). Composite scores were highly similar (correlation coefficient, 0.997) when calculated with versus without these procedures included, and they were retained in the study population to be consistent with current STS reporting conventions.

Composite Measure Components

The composite measure included a mortality domain and a morbidity domain. Rationale for focusing on these domains is discussed in Part 1 [1].

Mortality Domain

The mortality domain consisted of operative mortality. The STS-CHSD defines operative mortality as any death occurring in-hospital and any deaths occurring after discharge within 30 days of the operation. Further, inhospital deaths include deaths in the hospital performing the operation, in another acute care facility to which the patient is transferred, or in a chronic care facility up to 6 months after transfer.

Morbidity Domain

The morbidity domain consisted of major complications and postoperative length of stay (LOS).

MAJOR COMPLICATIONS—The postoperative complications included (described further in Part 1) are listed in Table 1 [1]. The STS-CHSD includes complications occurring during the same hospitalization as the operation or after discharge within 30 days of the operation. An “any or none” type of definition was used for the major complications measure; meaning, a patient with any one or more of the complications listed was counted as having a major complication. This approach can be thought of as analogous in concept to “freedom from major complications” for patients who did not experience any of the complications. Alternative approaches were also discussed (see Part 1), and it was agreed that weighting complications in relation to their effect on longer-term survival or quality of life would be desirable; however, adequate data in this area do not currently exist [1]. Efforts are underway to better capture these data moving forward [4].

Table 1 reports the frequency of each individual complication and the hospital-level correlation with the aggregate major complications rate. The correlation coefficients ranged

from 0.13 to 0.53, suggesting that no single complication dominates or explains all of the variation in the aggregate major complications end point.

LENGTH OF STAY—The rationale for including postoperative LOS in addition to major complications is described in Part 1 [1]. For analytic purposes, LOS was defined as 1 plus the number of days between the date of the operation and discharge, which facilitates modeling by ensuring LOS is never 0. To reduce sensitivity to outliers and satisfy modeling assumptions, LOS was transformed to the log scale, and LOS and values were capped at 90 days.

A sensitivity analysis was performed to understand how the inclusion of LOS might affect hospitals whose typical practice involves keeping patients undergoing the Norwood operation in the hospital until stage II (see Appendix). No hospital with a high proportion of such patients was classified differently with regard to composite measure performance whether LOS was included or excluded, suggesting that this practice is not likely to alter the overall assessment of quality using the composite measure at these hospitals and supporting retention of LOS in the final composite.

Statistical Modeling

Hospital-specific operative mortality rates, major complication rates, and distributions of LOS were estimated in a Bayesian multivariate hierarchical model (see Appendix). This methodology places more weight on a hospital's own case-mix adjusted end point when it is measured reliably and shrinks back toward a statistical prediction based on the model when the end point of interest is measured with greater sampling error (eg, hospitals with a small sample size) [5]. Another advantage of Bayesian approaches is that inferences about a hospital's performance are explicitly stated in probabilities, known as credible intervals (CrI). For example, based on a hospital's data, we might be 95% sure that its true performance is better than expected. Conventional *p* values and confidence intervals do not have a similar probability interpretation [6].

Different strategies were discussed, and we elected to model major complications and LOS only in survivors. This avoided "double counting" of events at the hospital level for a patient who, for example, had a major complication and died, and also created independent nonoverlapping domains for mortality and morbidity in the composite measure. However, we also recognized that it may be informative to hospitals to understand their rate of complications and LOS in the entire cohort versus survivors alone, and inclusion of this information from both cohorts in STS-CHSD feedback reports may be useful. Of note, there was a very high correlation between a hospital's complication rate overall versus survivors alone (0.98) and between a hospital's mean LOS overall versus survivors alone (0.99), suggesting benchmarking of quality was likely to be similar using either approach.

We further considered creating completely nonoverlapping domains and modeling LOS only in survivors without a major complication; however, this was felt to be clinically less desirable or intuitive, and there was high correlation (0.92) between hospital LOS estimates calculated in all survivors versus those without a major complication.

Case-Mix Adjustment

Operative mortality rates were adjusted for case-mix using the published STS-CHSD Mortality Risk Model [7]. The same factors were adjusted for when modeling the major complications and LOS end points, but coefficients were reestimated as described (see Appendix). STS-CHSD models are updated periodically, and we anticipate that future iterations of the composite score will use the most current version at the time of analysis.

Standardization of Measurement Scales

The construction of a composite measure based on different end points must account for the different measurement scales that apply to each end point [8]. For example, a hospital's average LOS as measured in days is not inherently commensurable with a mortality rate as measured by a binary proportion. To create a common measurement scale, hospital-specific performance for each end point was expressed in the form of risk-adjusted ratios (RARs). The RAR is the ratio of a hospital's true average results relative to the results that would be expected for a hypothetical average hospital with the same case-mix. The RAR produces a metric that accounts for case-mix and has a similar numerical interpretation for each outcome: mortality (RAR_{MORT}), major complications (RAR_{COMP}), and LOS (RAR_{LOS}).

Evaluation of Composite Measure Properties

We performed several analyses to explore the properties of candidate composite measures and individual component metrics. To verify that each individual component exhibits between-hospital variation, we evaluated the distribution of hospital-specific risk-adjusted mortality rates, major complications rates, and LOS. To assess clinical face validity (or the overall influence of the individual components on the composite score) and to verify that each of the individual component measures contributes statistical information but does not dominate the overall composite score, we calculated the hospital-level correlations between the point estimates for each individual component metric and the overall composite score.

To understand the ability of the composite measure to discriminate true hospital performance (signal) versus random statistical variation (noise), we estimated the composite measure's reliability and compared this value to the reliability of mortality alone. Reliability is a commonly used metric to assess suitability of performance measures and represents the signal-to-noise ratio or the proportion of variability in measured performance that can be explained by real differences in performance versus random statistical variation [9]. A reliability of 0.5 has been described as adequate or "moderate," and 0.7 is often considered "good" reliability [10].

Final Composite Calculation and Weighting

We considered several options for weighting the individual component metrics in the composite measure. Data are currently not available regarding patient or parent preferences for various health states; for example, preferences related to tradeoffs between certain morbidities and mortality. Similarly, information regarding the relative effect of specific morbidities on longer-term outcomes, such as quality of life or survival, is limited. Thus, in the absence of objective data to guide weighting, several types of commonly used mathematical formulas to combine individual component metrics were considered, with the

goal of maximizing clinical face validity such that in the final composite, the mortality domain would carry a greater influence on the composite score than the morbidity domain, and within the morbidity domain, major complications would have a greater influence on the composite score compared with LOS.

The final composite score was calculated as an equally weighted average of case-mix adjusted mortality and morbidity. Mathematically, the calculation was composite score = (mortality + morbidity)/2, where mortality = RAR_{MORT} and morbidity = $(RAR_{COMP} + RAR_{LOS})/2$. However, it is important to note that the “weights” do not necessarily equate to the importance of these individual components in the composite measure. This is because the RARs for the individual components have a different variance or distribution across hospitals. The relationship between individual components and the overall composite is better understood through assessing the correlation of each individual component metric with the overall composite, as described in the preceding section.

To further understand the influence of the individual component metrics on the overall composite and to ensure that complications and LOS did not have an undue influence, we estimated the effect on the composite score of changes in a hospital's complication rate and LoS versus mortality across several representative operations spanning the spectrum of case complexity: tetralogy of Fallot repair, arterial switch operation, and the Norwood operation (see Appendix).

Finally, we evaluated several other potential methods for calculating the composite score (see Appendix). For each alternative method, we assessed the composite measure properties described in the preceding section. Across methods, reliability and the proportion of hospitals classified as statistical performance outliers were generally similar. Therefore, we primarily considered clinical face validity and ease of interpretation in selection of the final weighting scheme, and the alternate methods were rejected because they resulted in greater influence of complications and LoS versus mortality on the overall composite score, which was less desirable.

Composite Score Performance Categories

Hospitals were classified as having better-than-expected performance as assessed by the composite measure if their 95% CrI for the composite score fell entirely below the STS average composite score, as having worse-than-expected performance if their 95% CrI for the composite score fell entirely above the STS average, and as having same-as-expected performance if their 95% CrI for the composite score overlapped the STS average.

Reporting Considerations

A recurring challenge in congenital heart surgery performance assessment is “noisy” or imprecise estimates due to relatively small sample sizes. This can be mitigated by aggregating data across multiple years to increase the sample size, but this makes the resulting performance information less timely. Alternatively, centers with very small sample sizes can be excluded, but this fails to provide feedback to low-volume centers. To explore these tradeoffs, we evaluated signal-to-noise ratio (reliability, as defined above) across

different analytic time frames and different hospital case-volume thresholds (see Appendix for details).

Results

Study Population Characteristics

A total of 78,425 operations from 100 hospitals were included. Average annual case volumes ranged from 14 to 923 across hospitals (median, 155 cases). In the overall cohort, the unadjusted operative mortality rate was 3.1%, the major complications rate was 13.3% (11.3% in survivors), and the median LoS was 7 days (interquartile range, 5 to 14 days; same in survivors). In subsequent sections, data are presented for major complications and LoS in the survivor cohort only.

Individual Metric Distributions and Correlation

Figure 1 displays the estimated distribution across hospitals of the individual metrics that comprise the composite measure. Each varied across hospitals: adjusted operative mortality rate (hospital median, 3.1%; 10th to 90th percentile, 2.1% to 4.4%), adjusted major complication rate (hospital median, 11.7%; 10th to 90th percentile, 6.4% to 17.4%), and adjusted median LOS (hospital median, 7.0 days; 10th to 90th percentile, 5.9 to 8.2 days).

The estimated Pearson correlations at the hospital level between individual component metrics were 0.23 (95% CrI, 0.03 to 0.43) for mortality versus major complications, 0.26 (95% CrI, 0.06 to 0.45) for mortality versus LOS, and 0.08 (95% CrI -0.03 to 0.19) for major complications versus LOS. These positive but relatively weak correlations suggest that although the individual component metrics may partially overlap, they likely represent somewhat distinct aspects of hospital performance that can inform the composite measure, and that one aspect of performance cannot necessarily be directly predicted from the others.

Composite Score Properties

CORRELATION OF INDIVIDUAL METRICS WITH THE COMPOSITE SCORE—We evaluated the hospital-level correlation between each individual component metric point estimate and the overall composite score. As summarized in Table 2, the Pearson correlation with the overall composite score was 0.87 for mortality, 0.70 for major complications, and 0.47 for LOS. These values support the clinical face validity of the composite measure and strategy outlined by the investigator team such that mortality carries the greatest influence on the composite, followed by major complications and then LOS. These data also suggest that all three individual metrics contribute information and that no single item dominates.

To further explore the influence of the individual component metrics on the overall composite score and to ensure that complications and LOS did not have an undue influence, we estimated the effect on the composite score of changes in a hospital's complication rate and LOS versus mortality. These calculations demonstrated that an absolute change of 1 percentage point in a hospital's adjusted mortality rate (eg, 4.1% vs 3.1%) would have the same effect on the composite score as an absolute change of 7.2 percentage points in a hospital's adjusted complication rate or a change of 4.5 days in a hospital's adjusted LOS.

We further examined these data across specific procedure types (tetralogy of Fallot repair, arterial switch operation, Norwood procedure; see Appendix), which further confirmed that LOS did not have too great an influence with the weighting scheme chosen.

RELIABILITY—The estimated reliability (signal-to-noise ratio) was 0.73 (95% CrI, 0.63 to 0.82) for the composite score compared with 0.59 (95% CrI, 0.45 to 0.70) for operative mortality alone, with a difference of 0.15 (95% CrI, 0.08 to 0.22; Table 2). This suggests that the composite measure had greater ability to discriminate true hospital performance compared with mortality alone.

HOSPITAL PERFORMANCE AS ASSESSED BY THE COMPOSITE MEASURE—Figure 2 displays the distribution of composite measure point estimates and 95% CrIs across hospitals. Overall, 16% had statistically better-than-expected performance, and 9% had statistically worse-than-expected performance (total of 25% classified as better or worse than expected). In contrast, when considering mortality alone, 11% of hospitals were classified as having statistically better-than-expected or worse-than-expected performance. The distribution of hospital performance across various categories of case volume and case-mix is summarized in the Appendix.

Hospital outcomes within composite performance categories are displayed in Figure 3. As expected, adjusted mortality, major complications, and LOS were all lowest in the group with better-than-expected composite performance. The magnitude of difference across performance categories was larger for mortality and major complications compared with LOS. This pattern provides additional internal validation that the composite scores accurately reflect the constructs they were designed to measure and that the weighting scheme functions as intended.

Reporting Considerations

We assessed the effect of different factors affecting sample size for reporting on reliability of the composite measure. Table 3 provides the estimated reliability for different potential reporting periods. As expected, reliability for the 4-year time frame is the highest (0.73) and declined with more narrow reporting windows (estimated reliability for 1-year time frame is 0.41). Table 3 also includes the estimated reliability associated with different hospital-level sample sizes. The estimated reliability reaches 0.5 when hospital case volumes are 120 or more per 4-year period (or >30 per year). If this threshold were adopted as a reporting cutoff, 8 hospitals in the current cohort would not have adequate reliability to be assigned a composite score.

Comment

This report describes the statistical methods and results related to development of the first pediatric and congenital heart surgery composite quality measure. The measure incorporates aspects of both morbidity and mortality and provides a more comprehensive view of quality compared with mortality alone. In addition, the composite measure has greater reliability compared with mortality and is better able to discriminate true hospital performance versus random statistical noise.

Similar to composite measures previously developed for adult cardiac surgery, we anticipate that this composite measure will be incorporated into STS-CHSD feedback reports to facilitate benchmarking and quality improvement activities [11]. Reports will include data regarding both the composite measure and its individual components. In the future, the composite measure may also be incorporated into the voluntary STS-CHSD public reporting initiative.

When interpreting the composite measure, it is important to recognize that the same important principals that apply to individual quality metrics apply in this case as well. Specifically, the composite metric can provide information regarding how a hospital is performing in relation to what would be expected for its particular case-mix (eg, a “rating”) [12]. It is not intended to be used to rank hospitals one against another, particularly those with differing case-mix. Currently, a hospital’s case-mix can be best understood by assessing the number and proportion of higher complexity cases performed (eg, STAT 4 and 5 cases). Such an approach, however, is related only to procedural case-mix and does not take into account potentially important patient factors. Ongoing work is focused on exploring a more comprehensive case-mix index.

Limitations

Although the statistical performance categories used in this report are one way to group centers, certain limitations apply to this and other methods of grouping centers into categories. Changes in performance categories may not necessarily always reflect a clinically meaningful change in performance, but rather small differences in certainty (>95% or <95%) about performance category assignment [6, 8]. It is also important to note that within the current methods for determining performance categories, most hospitals fall into the middle or “same-as-expected” category. Ongoing efforts are focused on exploring additional methods to continue to improve how performance data are conveyed to the public and other stakeholders [13].

Regarding case-mix adjustment, whereas we used the most current STS models, ongoing work is exploring updated methodology, and we anticipate that the methods for case-mix adjustment in this setting will continue to evolve and improve along with these efforts.

Finally, although the current composite measure provides a more comprehensive view of short-term performance, further efforts are needed to support more widespread collection of important longer-term disease-based rather than procedural-based outcomes data to continue to foster a more complete understanding of quality and improve our ability to assess outcomes most important to patients and families [4]. Such efforts could also inform more empiric methods for the inclusion and weighting of composite measure components (eg, certain complications) in the future.

Conclusion

We describe the development of the first composite quality metric in the field. This metric and future iterations can be used to inform ongoing efforts to better understand and improve pediatric and congenital heart surgery outcomes across hospitals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by funding from the National Heart, Lung, and Blood Institute (R01-HL-12226) to Dr Pasquali. Dr Pasquali also received support from the Janette Ferrantino Professorship.

References

1. Pasquali SK, Shahian DM, O'Brien SM, et al. Development of a congenital heart surgery composite quality metric: part 1-background and conceptual framework. *Ann Thorac Surg* 2018 9 15; [in press accepted manuscript].
2. Jacobs JP, Shahian DM, D'Agostino RS, et al. The STS National Database 2017 annual report. *Ann Thorac Surg* 2017;104:1774–81. [PubMed: 29153787]
3. Jacobs JP, Jacobs ML, Mavroudis C, Tchervenkov CI, Pasquali SK. STS-CHSD twenty sixth harvest. Durham, NC: STS and Duke Clinical Research Institute, Spring; 2017.
4. Pasquali SK, Ravishankar C, Romano JC, et al. Design and initial results of a programme for routine standardised longitudinal follow-up after congenital heart surgery. *Cardiol Young* 2016;26:1590–6. [PubMed: 28148316]
5. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg* 2012;255:703–7. [PubMed: 22388108]
6. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med* 1997;127:764–8. [PubMed: 9382395]
7. O'Brien SM, Jacobs JP, Pasquali SK, et al. The Society of Thoracic Surgeons Congenital Heart Surgery Database mortality risk model: part 1-statistical methodology. *Ann Thorac Surg* 2015;100:1054–62. [PubMed: 26245502]
8. Peterson ED, DeLong ER, Masoudi FA, et al. ACCF/AHA 2010 position statement on composite measures for healthcare performance assessment. *Circulation* 2010;121: 1780–91. [PubMed: 20351232]
9. Adams J The reliability of provider profiling. Available at http://www.rand.org/pubs/technical_reports/tr653.html. Accessed January 20, 2018.
10. Huffman KM, Cohen ME, Ko CY, Hall BL. A comprehensive evaluation of statistical reliability in ACS NSQIP profiling models. *Ann Surg* 2015;261:1108–13. [PubMed: 25211276]
11. Shahian DM, Edwards FH, Ferraris VA, et al. Quality measurement in adult cardiac surgery: part 1—conceptual framework and measure selection. *Ann Thorac Surg* 2007;83(Suppl):S3–12. [PubMed: 17383407]
12. Pasquali SK, Wallace AS, Gaynor JW, et al. Congenital heart surgery case mix across North American centers and impact on performance assessment. *Ann Thorac Surg* 2016;102: 1580–7. [PubMed: 27457827]
13. Pagel C, Jesper E, Thomas J, et al. Understanding children's heart surgery data: a cross-disciplinary approach to codevelop a website. *Ann Thorac Surg* 2017;104:342–52. [PubMed: 28318512]

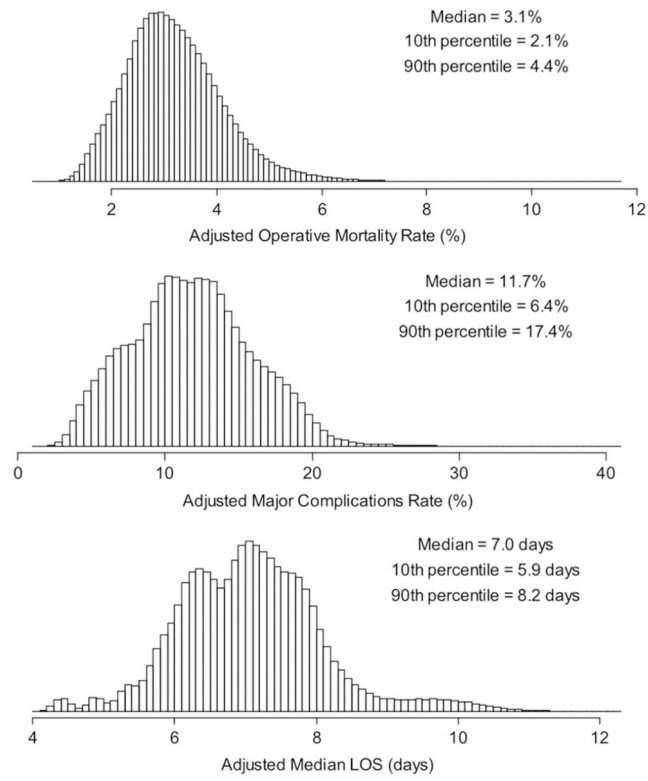


Fig 1. Distribution of adjusted mortality, major complications, and length of stay (LOS) across hospitals.

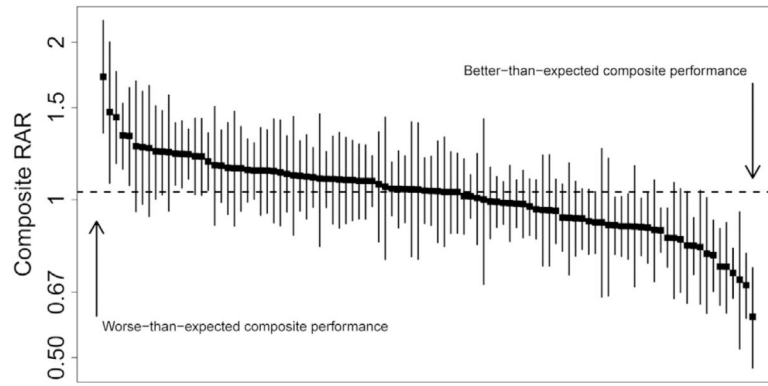


Fig 2. Composite measure point estimates and 95% credible interval (CrI) across hospitals. Hospitals are displayed in order of decreasing composite measure risk-adjusted ratio (RAR). Each black box represents a hospital's point estimate and the line represents the 95% CrI.

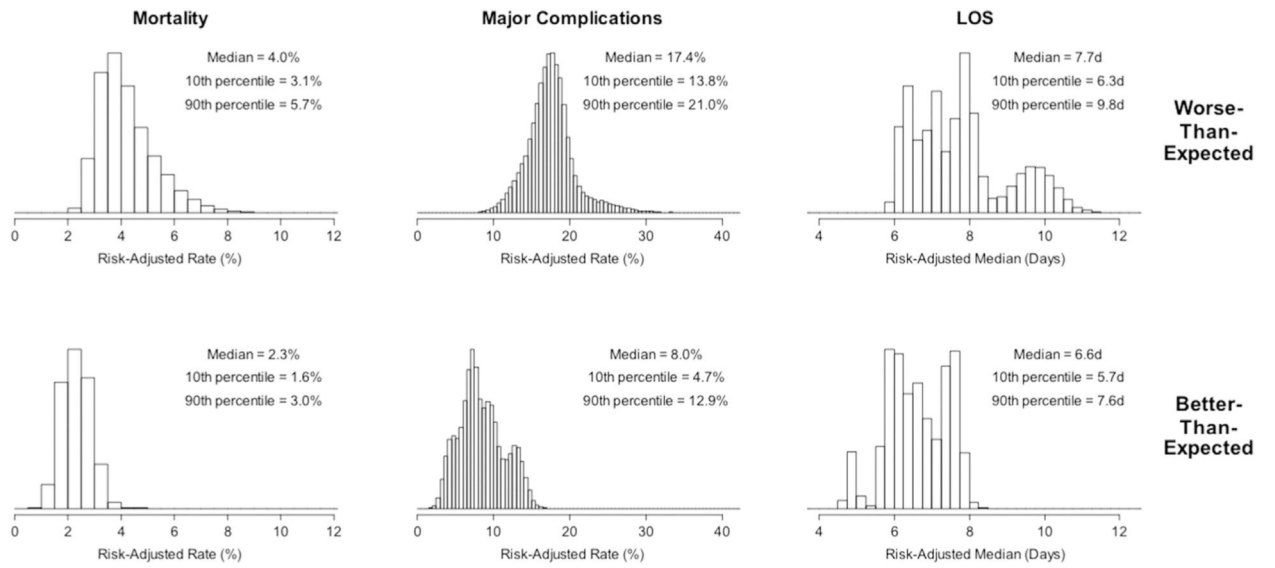


Fig 3. Summary data across composite measure performance categories. (LOS = length of stay.)

Table 1.

Individual Complications Comprising the Major Complications

Individual Complications	Events	Correlation With Major Complications ^a
	No. (%)	
Any major complication	8,553 (11.3)	1.00
Renal failure requiring		
Permanent dialysis	35 (0.0)	0.06
Temporary dialysis	289 (0.4)	0.17
Temporary hemofiltration	78 (0.1)	0.09
Neurologic deficit persisting at discharge	328 (0.4)	0.18
Arrhythmia requiring permanent pacemaker	997 (1.3)	0.32
Mechanical circulatory support	992 (1.3)	0.32
Phrenic nerve injury/ paralyzed diaphragm	768 (1.0)	0.28
Unplanned reintervention		
Reoperation for bleeding	1,215 (1.6)	0.36
Cardiac reoperation	2,399 (3.2)	0.51
Interventional cardiovascular procedure	1,530 (2.0)	0.40
Noncardiovascular procedure	2,519 (3.3)	0.52
Cardiac arrest	1,182 (1.6)	0.35

^aHospital-level correlation between the rate of each individual complication and the aggregate major complications rate.

Table 2.

Properties of the Composite Measure

Correlation of Individual Components With the Overall Composite Score	Pearson Correlation Coefficient
Operative mortality	0.87
Major complications	0.70
Length of stay	0.47
Signal-to-Noise Ratio	Reliability
Composite measure	0.73
Mortality alone	0.59

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Composite Measure Reliability Across Reporting Time Frames and Case Volumes

Reporting Time Frame	4 Years	3 Years	2 Years	1 Year
Composite reliability	0.73	0.67	0.58	0.41
Hospital Case Volume (4-Year Window)	50 Cases	120 Cases	200 Cases	300 Cases
Number of hospitals in current sample ^a meeting the specified case volume threshold	100	92	86	75
Composite reliability	0.30	0.50	0.63	0.71

^aN = 100 total.