

## PUBLIC INPUT SUMMARY REPORT

---

### **Project Title:**

*Hospital Quality Star Ratings on Hospital Compare*

### **Dates:**

The Call for Public Input period ran from August 30, 2017 to September 27, 2017.

The Public Input Summary was made on October 25, 2017.

### **Project Overview:**

The Centers for Medicare & Medicaid Services (CMS) has contracted with Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE) and Lantana, Inc. for both the development and maintenance of the publicly reported Star Ratings. The CORE contract name is Development, Reevaluation, and Implementation of Outcome/Efficiency Measures for Hospital and Eligible Clinicians, Option Year 3; contract number HHSM-500-2013-13018I, Task Order HHSM-500-T0001, and the Lantana contract is called Hospital Compare Support Contract (HCSC) Option Year 3; contract number HHSM-500-2013-13010I/HHSM-500-T0001. As part of the development and maintenance processes, CORE convenes groups of stakeholders and experts who contribute direction and thoughtful input during methodology refinement and maintenance.

The primary goal of this project is to develop a methodology for overall hospital quality star ratings to improve the usability and interpretability of *Hospital Compare* for patients and consumers. The Overall Hospital Quality Star Rating provides patients and consumers with a single measure to inform them about multiple dimensions of quality, represented by the existing measures on *Hospital Compare* and is capable of incorporating new measures that may be added in the future.

### **Purpose of Public Input Period**

This public input period focused on gathering feedback on the proposed methodology enhancements as well as areas of future work for the Overall Hospital Quality Star Rating from patients, healthcare consumers, and other stakeholders. Some commenters also provided additional comments that may apply to other aspects of the star ratings methodology or were beyond scope for the project. To maximize transparency and responsiveness, we have organized this report to first present comments related to the methodology enhancements and future work outlined, and subsequently included additional comments in a later section.

The previous two public input periods for this project focused on development of the Overall Hospital Quality Star Rating methodology.

### **Information About the Comments Received:**

Public input was solicited by email notifications to CMS listserv groups and web posts on the CMS Public Comment website (<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Public-Comments.html>).

CMS received comments from 34 commenters during the public input period. Specifically, we received comments from (see Table 1 in Appendix):

- Nine health systems
- Eight hospital associations
- One health care performance improvement company
- One healthcare leadership council
- Five hospitals
- Two medical universities
- One healthcare and policy foundation
- Six individuals

### **Stakeholder Comments—General and Project-Specific**

CMS received comments from the 34 commenters on various aspects of the Overall Star Rating project. Many of these comments addressed the requested topics for comment, while other comments were focused on other aspects of the methodology, or were considered beyond scope with respect to the Overall Star Rating. Comments focused on topics outlined in the public input period materials:

- Public Input Period Topics
  - Enhancements for the Overall Star Rating
    - Complete convergence
    - Removal of winsorization
    - Resequencing of reporting thresholds
    - Application of quadrature in latent variable models
  - Topics under ongoing reevaluation
    - Weighting of measure groups
    - Negative loadings
    - Public reporting thresholds
  - Future Areas of reevaluation
    - Stratification by hospital type or characteristics for peer group comparison
    - Measure inclusion
- Other comments received included input on the overall project, additional methodology topics, measure groups, star rating reporting, and stakeholder engagement.

### **Public Input Period Topics**

#### **1. Enhancements for the Overall Star Rating**

##### **1.1. Overall Enhancements**

Three commenters commended CMS for investigating and proposing improvements to the Overall Star Rating.

Nine commenters provided support for the enhancements outlined within the public input materials.

Seven comments noted specific improvements for the Overall Star Rating with the proposed enhancements.

- Four commenters specifically noted the broader distribution of star ratings.
- One commenter stated that the star ratings will be more informative for consumers with the broader distribution.
- One commenter noted the improved reliability, especially amongst the one- and five-star categories.
- One commenter conducted analyses based on the previous Star Rating SAS pack and proposed enhancements and found association, albeit less propounded than originally, between:
  - Number or type of measures and measure groups and hospital star ratings;
  - Select hospital and sociodemographic status (SDS) characteristics and the hospital star ratings.

**Response:** Thank you for your appreciation and support of the methodology enhancements for the Overall Star Rating. Just as individual measures and programs evolve, CMS will continue monitoring the utilization and interpretation of star ratings and will continue to investigate areas of improvement and solicit stakeholder input.

Four commenters had recommendations or suggestions for further changes beyond the recommended enhancements for the star ratings methodology.

- One commenter stated that, even with the enhancements, the Overall Star Rating does not provide meaningful information to consumers and providers.
- One commenter added that while the enhancements may well constitute marginal improvements on the existing methodology, these changes do not address underlying concerns.
- Two commenters stated that the proposed enhancements do not fix the underlying issues with the Overall Star Rating methodology.

Despite proposed enhancements and items for continued reevaluation, five commenters recommended removing or suspending the star ratings from *Hospital Compare*.

- One commenter recommended that CMS immediately suspend the star ratings until the method is improved and a process is put in place to include an independent audit to ensure the methodology is implemented correctly each time the methodology is changed and data is refreshed.
- Another commenter asked that CMS contract with independent outside experts to review the methodology and verify its accuracy.
- Another commenter said they had concerns with the star ratings approaches, measures included, and the lack of socioeconomic status (SES) adjustment in the ratings. The commenter recommended the removal of star ratings due to what they considered flaws in the methodology and unclear impact of the proposed methodology changes.
- One commenter said that hospitals are concerned about the Overall Hospital Quality Star Rating system, including issues with its methodology and doubt about whether it accurately reflects the quality of care provided by hospitals. The commenter urged CMS to suspend the star ratings and mitigate flaws in the system's measures and methodology.
- Another commenter expressed concerns regarding the methodology and recommended CMS remove the star ratings from *Hospital Compare*. The commenter recommended more time be taken with reviewing the stakeholder feedback. They added that they feel the methodology was

implemented with errors and they have concerns about potential consequences for patients that could result from painting an overly simplistic picture of hospital quality with the star rating system.

**Response:** CMS and the project developers appreciate the feedback regarding the star rating enhancements. The project developers have solicited different avenues of stakeholder feedback since the development of star ratings which includes public input periods, Technical Expert Panels, Provider Leadership Work Groups, and Patient and Patient Advocate Work Groups. The goal of the star ratings is to improve the usability and interpretability of aspects of hospital quality and quality measures for patients and consumers. Patient advocacy groups have supported the methodology and the reporting of star ratings. CMS will continue to evaluate stakeholder feedback to improve the methodology on an ongoing basis.

Two commenters noted analyses they completed regarding the enhancements.

- One commenter conducted analyses based on the previous Star Rating SAS pack and proposed enhancements and found that the methodology, with enhancements, still demonstrates bias toward smaller hospitals serving more advantaged populations.
- One commenter conducted analyses on the proposed enhancements using the publicly available December 2016 SAS pack and found that academic medical centers shift more to the one- and two-star categories.

Three commenters noted concerns with the methodology enhancements.

- One commenter stated that the enhancements create additional barriers to patient understanding of hospital care.
- One commenter expressed concern that the enhancements result in a fourfold increase in hospitals receiving a five-star rating with twice as many five-star hospitals than one-star hospitals, which could be seen as lowering expectations. The commenter suggested that CMS conduct analyses to determine the level of performance hospitals are attaining to receive each star rating category with the proposed enhancements.
- One commenter suggested conducting further analyses on each separate enhancement and to determine any impact of the proposed enhancements on hospital types.

**Response:** The goal of the Overall Star Rating is to summarize the existing quality information on *Hospital Compare* in a way that is easy to use and understand for consumers. We will continue to test the impact of the methodology enhancements and will continue to investigate areas of improvement, including stratification and reporting thresholds as a means to compare similar hospitals. While the totality of the enhancements does result in a shifted star ratings distribution, CMS acknowledges that no inherent, consensus benchmarks exist to distinguish between star ratings categories. As such, the current methodology seeks to utilize clustering to avoid the arbitrary assignment of star ratings by an approach such as quintiles, and instead accommodates shifting distributions of performance.

One commenter had questions about the proposed enhancements and impact on the Overall Star Rating. The commenter asked how extreme outliers were identified and handled and which statistical program was used to run the k-means clustering analysis. The commenter explained that statistical program and procedure to run the calculation may make a difference and influence the calculation and distribution.

Four commenters requested that CMS provide additional data and model output statistics for better transparency, stakeholder recommendations, and interpretation of the methodological impact. Some of those commenters had specific recommendations.

- Two commenters requested that CMS specifically provide statistical results, such as R-square, Pseudo F, CCC statistic, ANOVA, etc.

**Response:** CMS has conducted a variety of relevant model testing statistics related to the methodology enhancements. As future analyses are conducted, CMS will seek to include model output statistics in the publicly posted Quarterly Updates and Specifications Report.

Previously, winsorization of hospital summary scores was used to limit the effect of extreme outliers, k-means clustering was only run using one iteration, and the public reporting thresholds were applied after k-means clustering or the assignment of hospitals to star rating categories. Applying the reporting thresholds before k-means clustering and running k-means clustering to complete convergence results in a broader distribution of star ratings with fewer outliers making the use of winsorization no longer necessary. In SAS, PROC FASTCLUS is used to run k-means clustering. The option of 'converge =0' is used for running to complete convergence and the option of 'maxiter = 1000' is used to define the maximum number of iterations. During testing, k-means clustering required approximately 15-30 iterations to reach convergence in prior reporting periods.

### 1.2. K-Means Cluster to Complete Convergence

Eleven commenters expressed specific support for and provided comments regarding running k-means clustering to complete convergence. Some of those commenters had specific comments related to complete convergence.

- Three commenters noted that multiple iterations are required to achieve convergence and categorization of hospitals that are more similar to other hospitals within a cluster than to hospitals in other clusters.
- One commenter stated that complete convergence would improve the stability and reliability of k-means cluster or star rating assignment.

One commenter had questions about running k-means clustering to complete convergence. The commenter asked how CMS assigned hospitals to particular clusters without complete convergence and whether the variation among clusters was statistically significant.

One commenter expressed concern about running k-means clustering to complete convergence. The commenter acknowledged the high reliability for one- and five-star categories but expressed concerns for the moderate reliability of the two-, three-, and four-star categories.

**Response:** Thank you for your support and comments. CMS believes that calculation of clustering to complete convergence will result in increased stability and reliability of star rating assignments. In the past, the default SAS setting of one iteration was used for the k-means clustering analysis, which still produced five clusters of hospitals or star rating categories with five-star hospitals statistically better than four-star hospitals, four-star hospitals statistically better than three-star hospitals, and so on. Complete convergence ensures that the k-means clustering analysis runs until a more stable set of clusters is established. In other words, k-means clustering is run multiple times until hospitals are less

likely to be classified into different star rating categories. CMS believes this further refines the clustering method to be more stable than the default approach initially used.

### 1.3. Winsorization Removal

Seven commenters agreed that running k-means clustering to complete convergence obviates the need for winsorization. Some of those commenters provided specific comments about winsorization.

- One commenter noted that the use of winsorization may mask true variability and distribution of hospital summary scores.

Instead of removing winsorization, one commenter suggested widening the summary score thresholds by the 0.25<sup>th</sup> and 99.75<sup>th</sup> percentiles, for example.

**Response:** Thank you for your support for the removal of winsorization for hospital summary scores. Winsorization of hospital summary scores was a policy-decision based on stakeholder and consumer requests for a broader distribution of hospitals. The totality of proposed enhancements achieves a broad distribution on its own, making winsorization no longer necessary.

### 1.4. Resequencing of the Reporting Threshold

Thirteen commenters expressed specific support for resequencing the reporting thresholds so that only hospitals that ultimately receive a star rating are included in the k-means clustering calculation. Some of those commenters provided further comments about their support for resequencing of the reporting thresholds.

- Eight commenters noted that applying the reporting thresholds prior to k-means clustering will lessen the influence of hospitals that do not report enough measures or measure groups on the clustering parameters.
  - One commenter expressed further concern for the inclusion of low volume hospitals within the k-means clustering analysis for star rating assignment. The commenter stated that low volume hospitals perform exceptionally well or poor on measures and the Overall Star Rating, forcing large hospitals to appear average.
  - One commenter conducted analyses and found that the lower number of measures a hospital reports, the more likely they will receive a five-star rating.
  - One commenter conducted analyses and found that hospitals that do not meet the reporting thresholds, on average, have higher summary score than hospitals that meeting the reporting threshold.
- One commenter stated that resequencing the reporting thresholds so only hospitals that meet the requirements are included in the star rating assignment increases transparency.
- One commenter noted that resequencing the reporting thresholds so only hospitals that meet the requirements are included in the star rating assignment results in a broader distribution and more one- and five-star hospitals.

One commenter expressed concerns about resequencing the reporting thresholds so that only hospitals that ultimately receive a star rating are included in the k-means clustering calculation. The commenter stated that the resequencing of reporting thresholds is not helpful to healthcare consumers who may not understand the need or significance of requiring hospitals to meet public reporting thresholds.

Five commenters suggested applying the reporting threshold before latent variable modeling of the measure groups so ineligible hospitals do not impact measure group scores. Some of those commenters had specific comments about applying the reporting threshold before latent variable modeling.

- One commenter conducted sensitivity analyses of the latent variable models and found including hospitals that don't meet the reporting thresholds affects measure group scores as well as the cluster parameters.

**Response:** Thank you for your support of and suggestions for resequencing the public reporting thresholds. We also believe that only including hospitals that will ultimately receive a star rating in the k-means clustering analysis will provide better comparisons among eligible hospitals. The reporting threshold is applied after the latent variable modeling, and before clustering, in order to include as much information in latent variable modeling as possible which helps model performance without having significant impact in later steps. To note, any time a parameter of the underlying data is changed, a change in the distribution should be expected. Reporting thresholds ensure that consumers are using reliable and meaningful data to make decisions. Additional information about measure-level and the Overall Star Rating thresholds are available to stakeholders and consumers that may be interested.

### 1.5. Quadrature

Seven commenters expressed support for the application of iterative adaptive quadrature. Some of those commenters had specific comments about their support for adaptive quadrature.

- Two commenters provided support based on concept and theory that it would result in more stable estimates. However, the commenters cannot provide feedback based on a full evaluation without analytic exhibits.
- Two commenters expressed appreciation for testing of iterative adaptive quadrature in prior reporting periods and the potential improvements in stability of measure groups scores and star ratings.
- One commenter stated that the application of iterative adaptive quadrature would increase model convergence.

Two commenters made suggestions regarding the application of iterative adaptive quadrature within latent variable modeling for the Overall Star Rating.

- One commenter suggested 1) updating and elaborating the model specifications related to quadrature, including defining all the variables; 2) posting the proposed SAS code with explanation of the mathematical model with algorithms; and 3) documenting each SAS function call and explaining the parameters.
  - The commenter stated that documenting the model and SAS code would clarify what the SAS code is computing, allow stakeholders to comment on the appropriate approximation technique, and allow hospitals to better understand how hospitals are being rated and how hospitals can improve.
- One commenter suggested that CMS increase transparency by making the adaptive quadrature testing public and reviewing the application of the iterative adaptive quadrature with the Technical Expert Panel to allow for more informed and meaningful stakeholder contributions.

**Response:** Thank you for your support of and comments for the application of iterative adaptive quadrature. The details and results of iterative adaptive quadrature will be available within the December Updates and Specifications Report, December SAS Pack and Documentation, and the updated

Comprehensive Methodology Report (v3.0) on the QualityNet webpage. Questions and comments can be sent to the Hospital Quality Star Ratings inbox at [cmsstarratings@lantanagroup.com](mailto:cmsstarratings@lantanagroup.com).

## 2. Ongoing Reevaluation Topic: Weighting of Measure Groups

### 2.1. Process and Efficiency Group Weighting

Six commenters made suggestions for the weighting specific to the process (Effectiveness of Care and Timeliness of Care) and efficiency (Efficient Use of Medical Imaging) measure groups. Some comments contained multiple points shared by others so comments are broken apart further and grouped below. The current measure group weighting for the Overall Star Rating is included in [Table 1](#) of [Appendix A](#).

- Five commenters favored an increase in weighting for process measure groups.
  - One commenter stated that process measures are important since they assure safe and proper care.
  - One commenter stated that retaining lower weights for process measures insinuates that process measures are relatively inconsequential within the context of hospital quality improvement efforts, which is inaccurate and unhelpful in aiding the understanding of healthcare information for consumers.
  - One commenter stated that process measure data is updated quarterly and allows for timely hospital improvements.
- One commenter suggested retaining or implementing a similar weight for process measure groups. The commenter proposed 5% weight for process measure groups, which is between the 8% included in the alternative approach and no weight given by the Hospital Value-Based Purchasing (HVBP) program.
- One commenter suggested that the Effectiveness of Care process measure group should be weighted more heavily than the Efficient Use of Medical Imaging measure group based on importance.
- One commenter favored an increase for the Efficient Use of Medical Imaging measure group, stating that efficiency measures are important considering health risks of radiation and financial implications associated with overuse of imaging.

### 2.2. Outcome Group Weighting

Nine commenters made suggestions specific to the outcome (Mortality, Safety of Care, and Readmission) and Patient Experience measure groups. Some comments contained multiple points shared by others so comments are broken apart further and grouped below. The current measure group weighting for the Overall Star Rating is included in [Table 1](#) of [Appendix A](#).

- Three commenters recommended retaining higher weights for outcome measures and patient experience.
  - One commenter stated that Mortality and Patient Experience measures are most important since they reflect a patient's life and experiences. Safety of Care and Readmission are also importance because they reflect avoidance of adverse events and readmissions.
  - One commenter stated that hospital performance on outcome measures is highly associated with quality of outpatient and emergency care through establishing processes that focus on effective and efficient care.
  - One commenter noted that outcome measures are meaningful to patients and usually have sufficient sample size, number of events, and robust risk adjustment.

- Four commenters suggested increasing the weight for the Mortality measure group.
  - Two commenters stated that mortality is the most important aspect of care to consumers for many conditions and procedures.
  - One commenter noted that mortality is the measure group currently least correlated with the Overall Star Rating.
  - One commenter stated that the additional time taken to collect and process mortality data is worth the effort and results in valuable information for patients.
- One commenter suggested lower weights for Mortality, Safety of Care, and Readmission.
- Two commenters suggested lower weight for Safety of Care, considering the two-year delay in reporting performance data for most measures, such as the Agency for Healthcare Research and Quality (AHRQ) Patient Safety Indicator composite measure (PSI-90) and Hip/Knee Complications. The measures do not reflect current performance to consumers or actionable data for hospitals.
  - However, the commenter supports a higher weight for Safety of Care than Mortality and Readmission because of the timely and clinically relevant National Healthcare Safety Network (NHSN) measure.
- One commenter suggested lower weights for Safety of Care and Readmission measure groups given concerns with validity and lack of risk adjustment for social determinants of health for the measures within these measure groups.
- One commenter advocated for a lesser weight for Readmission and Patient Experience.
- One commenter suggested a 20% or lower weight for Safety of Care because of the domination of the PSI-90 composite measure, which assigns average scores to, rather than excluding, low volume hospitals through the reliability adjustment.
- One commenter suggested reducing the weight of the Safety of Care domain due to concerns about the reliability of data collection and inadequate risk adjustment. In addition, there is variation in surveillance methods for hospital-acquired infections, complicating consumers' ability to reliably compare infection outcomes across hospitals. The infection measures do not adequately account for differences in clinical characteristics, patient acuity, or SDS factors, resulting in high standardized ratios for hospitals caring for more complex patients. The commenter would support increasing the Safety of Care weight should these methodological concerns be addressed.
- Three comments made suggestions specific to Patient Experience.
  - One commenter suggested a 20% or lower weight for Patient Experience because it is most correlated with the Overall Star Rating, likely because of the additional weighting it receives when hospitals don't report other measure groups. In addition, the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey may be inherently biased against safety net hospitals since it accounts for patient mix but not hospital-level disparity in financial resources between safety net hospitals and hospitals that serve more advantaged and insured patients. This resource disparity is beyond hospitals' control and causes disparities in capital infrastructure and staffing levels, influencing patient experience.
  - Two commenters suggested higher weights for patient experience.
    - One commenter noted that the data is updated quarterly and allows for timely hospital improvements.
    - One commenter stated that patient experience is an important indicator of quality and believes that the data is collected with high fidelity and reliability.

- One commenter suggested decreasing the weights of outcome measure groups stating that process measure groups that influence outcomes are under the control and responsibility of a hospital while there are too many variables to consider for outcome measures, some of which are out of the hospitals' control.

### 2.3. Alternative Weighting Schemes

Ten commenters communicated preferences if choosing between the current measure group weighting scheme and the alternative examples presented in the public input period materials. The current measure group weighting scheme ([Table 1](#)) and example alternative measure group weighting schemes ([Table 2](#)) are included in [Appendix A](#).

- Four commenters expressed a preference for alternative #1.
  - One commenter based their preference on fair distribution of weighting and measure importance.
  - One commenter prefers alternative #1 because they believe Safety of Care should have a greater or equal weight to Patient Experience since safety is much more objective and less easily affected by marking strategies.
  - One commenter based their preference on an emphasis for patient safety and effectiveness of care and the decrease in weight for the Efficient Use of Medical Imaging measure group, which the commenter does not find relevant or useful.
- Five commenters expressed a preference for alternative #2.
  - One commenter recommends alternative #2 because of concerns regarding the validity and absence of adjustment for social determinants for measures within the Safety of Care and Readmission measure groups.
- Two commenters suggested retaining the current measure group weights.

Eight commenters proposed additional alternative measure group weighting schemes. Some comments contained multiple points shared by others so comments are broken apart further and grouped below.

- One commenter proposed an alternative measure group weighting scheme with 20% weight for Mortality and Patient Experience, 15% weight for Safety of Care and Readmission, and 10% weight for Effectiveness of Care, Timeliness of Care, and Efficient Use of Medical Imaging.
- One commenter proposed an alternative measure group weighting scheme that separated measure groups into categories for Quality, Efficiency, and Patient Experience.
  - Within the Quality category, Mortality was assigned a 30% weight, Safety of Care a 20% weight, and Effectiveness of Care and Timeliness of Care a 5% weight for a subtotal of 60%.
  - Within the Efficiency category, Readmission was assigned a 15% weight and Efficient Use of Medical Imaging a 5% weight for a subtotal of 20%.
    - The commenter categorized readmission within the efficiency category because of the 30-day timeframe. In addition, readmission measures could be seen as valid measures of efficient use of the inpatient setting versus the outpatient or ambulatory setting or valid measure of quality, access, and care coordination in the community.
  - The Patient Experience category and measure group was assigned a 20% weight.
- One commenter suggested equal total weights for outcome and process measure groups, similar to the approach used for the Leapfrog composite.
  - Mortality, Safety of Care, and Readmission would each be weighted as 16.6% for a total of 50%.

- Patient Experience, Effectiveness of Care, and Timeliness of Care would each be weighted 15% and Efficient Use of Medical Imaging would be weighted 5% for a total of 50%.
- One commenter suggested an alternative measure group weighting scheme of 25% for Mortality, 20% for Patient Experience, 15% for Safety of Care and Readmission, 10% for Effectiveness of Care and Timeliness of Care, and 5% for Efficient Use of Medical Imaging.
- Three commenters suggested a weighting approach with heavier weights for groups that reflect clinical-specific measures and more current performance for consumers and more actionable data for hospitals.
  - One comment suggested the alternative weighting scheme would assign 20% weight to Safety of Care and Patient Experience, 18% to Mortality and Readmission, 10% weight to Effectiveness of Care and Timeliness of Care, and 4% weight to Efficient Use of Medical Imaging.
  - One commenter further suggested that the measure group weights align with other CMS programs, such as HVBP.
  - One commenter suggested assigning measure weight within each measure group based on timeliness of the data with a higher weight for more currently available data and lower weight for older data and clinical relevance with a higher weight for NHSN and lower weight for PSI-90 and Hip/Knee Complications. From there, determine percentage of data completeness. For example, hospitals must report on at least 70% of measures within a measure group across five of the seven measure groups to receive a star rating.
- One commenter proposed an alternative measure group weighting scheme with 26% weight for Mortality and Patient Experience, 18% weight for Safety of Care and Readmission, and 4% weight for Effectiveness of Care, Timeliness of Care, and Efficient Use of Medical Imaging.
- One commenter proposed an alternative measure group weighting scheme based on integrity of measures and importance of the measure group in determining a hospitals' performance. with 30% weight for Mortality, 22% weight for Patient Experience, 18% weight for Safety of Care and Readmission, and 4% weight for Effectiveness of Care, Timeliness of Care, and Efficient Use of Medical Imaging.

One commenter stated that the weighting of measure groups should be based on the reliability and validity of the measures and measure groups, and expressed concern with the use of latent variable modeling to establish measure groups.

**Response:** CMS thanks you for your comments and suggestions for measure group weightings and alternative approaches. The current weights are based on group importance, consistency with existing CMS policies and procedures, and stakeholder input. The weights are policy-based and modified from the HVBP program. As noted by the breadth of comments and suggestions, there are many differing options regarding the weighting scheme. CMS will continue to evaluate this topic and compile feedback from stakeholders for future improvement.

#### 2.4. Redistribution of Measure Group Weights

One commenter expressed concerns about the redistribution of weights when a hospital does not have measures in one or more measure groups, which could increase the weight for a measure group that is reflecting performance on less than two measures.

- The commenter further explained that past data shows hospitals with only three reported measure groups need an above average performance in one measure group to receive a five-

star rating while hospitals with all measure groups need an above average performance in at least three measure groups to receive the same rating. In addition, hospitals that report all measure groups have high or equal performance to hospitals that report only three measure groups with a lower star rating. The commenter believes this is due to the redistribution of weights and low thresholds.

- The commenter suggested redistributing the weight of measure groups for which a hospital does not report to measure groups with at least three measures while measure groups with less than two measures keep the standard weight.

**Response:** Thank you for your suggestions regarding redistributing measure groups weights. CMS has evaluated the effect of number of measures and number of measure groups on the final Overall Star Rating and found weak correlation of mixed direction between the number of reported measures within a group and the group score. These results do not indicate systematic biases that would disadvantage hospitals based on the numbers of measures reported. CMS will continue to evaluate the topic and request stakeholder input for future improvement to the redistribution of weights.

Four commenters communicated concerns and recommendations about measure group weighting decisions.

- One commenter expressed concerns about the conflicting suggestions for measure group weighting based on stakeholder input from the Technical Expert Panel, measures currently available on *Hospital Compare*, and the Overall Star Rating guiding principle to align with other CMS programs.
  - For example, *Hospital Compare* includes many process measures and some Technical Expert Panels favor increased weights for process measure groups, yet performance-based payment programs will not include process measures as of fiscal year 2018.
  - In addition, CORE and the Technical Expert Panel have identified Medicare spending per beneficiary measures as non-directional but CMS has given them 25% weighting within the HVBP scoring program as of fiscal year 2018.
- Two commenters expressed concern that weighting of measure groups have significant influence on the Overall Star Rating. The commenter suggested establishing measure group weights based on a clear analytical approach and survey of patient and caregiver preferences that would be helpful in making decisions about care.
- One commenter encouraged continued engagement with the Technical Expert Panel and public while considering aspects of quality that hospitals have control over as well as what measures incentivize the best patient care.

**Response:** Thank you for your comments, suggestions, and expressed concerns. One of the criteria considered for measure group weights is the consistency with existing CMS policies and priorities as alignment across programs is a priority. We also appreciate the suggestions to continue exploring analytic approaches to establish weights and continue to compile stakeholder feedback. CMS will continue evaluating measure group weighting for future improvement.

## 2.5. Negative Loadings

Eight commenters expressed concern that hospitals that perform poorly on measures with negative loadings are receiving higher measure group scores and performance categories than hospitals performing well on the same measures. Some of those commenters noted specific concerns about the impact of negative loadings.

- A commenter provided a scenario in which a hospital received “worse” than the national average for Readmission and Safety of Care while other hospitals received “worse” than the national average for Readmission, Safety of Care, Patient Experience, and Timeliness of Care but received a higher star rating.
- One commenter noted that if the measures do not have much impact, and the impact they do have is the opposite of what was intended, then there is no reason to keep the measures in the star rating.
- One commenter expressed concern that negative loadings discourage hospitals by penalizing them for performing well on these measures.
- Two commenters pointed out that the theory that a measure with a negative loading would adversely affect a hospital’s star rating is counterintuitive to the project goals.
- One commenter pointed out that this potential effect would be misleading to consumers.

One commenter proposed that negative loadings are problematic with respect to face validity and could be occurring in measures where high-quality hospitals perform poorly due to circumstances. For example, a great hospital could be located next to a nursing hospital and have a higher mortality rate than a low-quality hospital that is located next to a college campus.

Four commenters noted concerns about the Healthcare-Associated Infections Clostridium Difficile measure (HAI-6) having a negative loading.

- Two commenters noted CMS efforts to prevent clostridium difficile in the hospital setting through programs, such as the Hospital Acquired Conditions (HAC) Reduction Program, are tied to financial incentives. However, hospitals could receive a lower star rating for performing well on the measure.
- One commenter stated that it is implausible that an infection associated with poor hand hygiene is a sign of quality.
- One commenter noted that the negative loading of HAI-6 is likely due to the high loading of the PSI-90 measure within the Safety of Care measure group and therefore, PSI-90 should be removed from the Overall Star Rating.

Seven commenters recommended that CMS remove measures with negative loadings. Some commenter provided further input on their recommendation to remove measures with negative loadings.

- One commenter noted that, regardless of the impact on the Overall Star Ratings is relatively modest, including measures with negative loadings is not easily understood by patients, caregivers, and the public and does little to inform overall patient care and public reporting.
- One commenter recommended removing measures that have results that are not statistically significant at 5% because hospitals will not have prior knowledge regarding which measures are negatively loaded with the group score.

Three commenters recommended that, if measures with negative loadings are removed, only measures with statistically significant ( $p < 0.05$ ) should be removed.

Three commenters recommended that, if measures with negative loadings or positive but statistically nonsignificant or near-zero loadings are removed, CMS reassesses all measures each iteration of the Overall Star Rating since the loadings may fluctuate over time as hospital performance evolves.

Three commenters provided suggestions for measures with consistently negative or near zero and non-significant loadings.

- One commenter recommended evaluating the appropriateness of including measures with consistently negative loadings.
- One commenter recommended removing measures with consistently low, near zero and non-significant loadings, should the Overall Star Rating continue to utilize latent variable modeling.
- One commenter suggested measures with negative loading two quarters in a row be permanently removed from the rating.

Five commenters recommended including all measures with negative loadings, even those that have statistically significantly negative loadings, and positive but statistically nonsignificant or near-zero loadings.

- Two commenters stated that retaining all measures would ensure technical consistency and prevent any effect on the loading direction of other measures.
  - One of those commenters noted that negative loadings may be unavoidable and retaining all measures with negative loadings and positive but statistically nonsignificant or near-zero would at least minimize the number of measures with negative loadings.
- One commenter does not believe measures should be excluded based on directionality and negative loadings are unlikely to affect the Overall Star Rating.
- One commenter noted that the impact of excluding measures with negative loadings on other measures is unknown.
- One commenter noted that there is little empiric evidence that improvement on a measure with a negative loading would lead to a lower overall rating. In addition, removing and including measures based on their loading would result in inconsistent measure inclusion and could confuse consumers and the public, which would counter the goals of the Overall Star Rating.

One commenter deferred comments until they can investigate the effect of negative loadings, or the exclusion of negative loadings, and positive but statistically nonsignificant or near-zero loadings on measure group scores with the December 2017 dataset. Until that time, the commenter recommends that CMS holds off on making decisions regarding negative loadings.

**Response:** Thank you for your comments and suggestions. CMS will consider your recommendations. Please note that no measure included in the Overall Star Rating to date has had a statistically significantly negative loading and there is no empirical evidence to date that hospitals are receiving lower star ratings based on good performance for measures with negative loadings. CMS hopes to utilize this public input to inform future decision making including any potential modifications to the methodology with regards to negative loadings. CMS welcomes your review of the December SAS pack and results. The December data and results will be made available through the posting of the SAS pack and input file on QualityNet.

## 2.6. Public Reporting Thresholds

Five commenters relayed that the current public reporting thresholds (three measures; three measure groups with at least one outcome measure group) should be modified.

- One of those commenters suggested reducing the number of measure groups required to receive a star rating so smaller hospitals can consistently receive a star rating.

- One of those commenters suggested increasing the number of measures required to receive a star rating for more homogeneous comparison of eligible hospitals since the number and type of measure reporting can vary by hospital size, hospital type, and patient populations served.
- One commenter agreed with the need for measure thresholds within measure groups but suggested using a percentage of measures for which data are available considering the varying number of measures within each measure group.
- One commenter suggested assigning individual measure weights within each measure group based on timeliness of the data and clinical relevance. For example, measures with more current data would have higher weights and measures with older data would have lower weights. From there, a hospital would need 70% of the measure group weight in at least five measure groups to receive a star rating.
- One commenter agreed with requiring three measures and three measure groups to receive a star rating but suggested requiring at least two outcome measure groups, instead of only one, to limit star ratings for hospitals that have sufficient volume for comparison. The same commenter asked that this approach be evaluated prior to implementation to ensure no hospitals are unfairly disqualified from star ratings.

Instead of modifying the reporting thresholds, two commenters suggested adjusting the summary score performance based on the volume of measures reported.

Six commenters suggested no change to the number of measures or measure groups required to receive a star rating.

**Response:** Thank you for your suggestions. Please note that the Overall Star Rating is meant to summarize the existing measures on *Hospital Compare* and is not meant to evaluate individual measures including specific data collection periods.

During the initial development of the Overall Star Rating, analyses were conducted to determine the number of measures and measure groups needed to maximize the number of hospitals receiving a star rating while maintaining reliability. The results of those analyses concluded that three measures within a measure group and three measure groups with at least one outcome group allowed 78% of hospitals in the nation to receive a star rating with a reasonable reliability score. Based on the comments and suggestions received during the public input period, the development team will continue to test reporting threshold options.

Despite stakeholder analyses from last year that claim hospitals reporting fewer measures receive higher star ratings, the development team has found little correlation and no systematic relationship between the number of measures reported and measure group scores or star ratings in more recent reporting periods.<sup>1</sup> However, distinguishing between hospital performance based on number and type of measures reported is an ongoing discussion for which CMS is considering peer group comparison, or stratification. CMS will continue to engage stakeholders on this topic in the future.

## 2.7. Measures Included After Public Reporting Thresholds

---

<sup>1</sup> Venkatesh, A., Qin L., Lin H., et al. Relationship between hospital quality measure reporting and Overall Hospital Quality Star Ratings performance. Poster presented at: Academy Health; June 25-27, 2017; New Orleans, LA.

Five commenters provided input on whether the current inclusion of all other measures once the reporting threshold is met should be modified.

- Two of those commenters suggested not including any additional measures or measure groups in the star rating calculation once the reporting thresholds are met.
- One of those commenters suggested including additional measure groups with at least two measures once the threshold is met with the goal of maximizing the number of hospitals receiving a star rating while decreasing the chance of a single measure having a large impact on a measure group or the star rating.
- Two of those commenters suggested retaining the additional measures in the star rating calculation once the reporting thresholds are met.

**Response:** Thank you for your suggestions. The original decision to include additional measures and measure groups once the reporting thresholds are met was based on the principle of inclusivity of available measures and information. CMS will investigate further and consider each of your suggestions for potential testing.

One commenter deferred comments until the December 2017 data and SAS pack are released for testing.

**Response:** CMS strives to maintain transparency and welcomes further input. The December 2017 data and SAS pack will be available on the QualityNet website during December reporting.

### 3. Future Areas of Reevaluation

#### 3.1. Stratification

Fourteen commenters suggested stratifying the Overall Star Rating. Most commenters suggested multiple approaches to stratification. Some comments contained multiple points shared by others so comments are broken apart further and grouped below.

- Five of those commenters suggested stratifying by specialty hospitals versus general hospitals.
  - One commenter further suggested calculating star ratings for general hospitals, defined as hospitals having a publicly reported mortality rate for heart failure and pneumonia, based only on general hospital data and calculating star ratings for specialty hospitals based on all hospital data.
- Four of those commenters suggested stratifying by hospital size or number of beds.
- Two of those commenters suggested stratifying by annual discharge volume.
- Six of those commenters suggested stratifying by academic hospitals or teaching status versus community hospitals.
- Four of those commenters suggested stratifying by location, such as urbanicity and rurality.
- Five of those commenters suggested stratifying by services offered.
  - Two of the commenters specifically suggested stratifying by trauma service.
  - Two of the commenters specifically suggested stratifying by bone marrow transplant service.
  - Two of the commenters specifically suggested stratifying by all solid organ transplant service.
- Two of those commenters suggested stratifying by SDS characteristics or patient demographics.
- Three of those commenters suggested stratifying by completeness of reported measures, number and type of measures reported, or number of measure groups reported.

- One of those commenters suggested stratifying by acute versus long-term care.
- One of those commenters suggested stratifying by health care systems versus independent hospitals.
- Two commenters suggested stratifying by total outpatient visits.
- Two commenters suggested stratifying by acute transfers into the hospital.
- Three commenters suggested stratifying by case mix index.
- Two commenters suggested stratifying by inpatient surgical cases as a percentage of all admissions.
- Two commenters suggested stratifying by outpatient surgical cases as a percentage of total surgical cases.
- One commenter suggested stratifying by critical access designation.
- One commenter suggested stratifying by disproportionate hospital share patient (DSH) percentage.

**Response:** Thank you for your comments. CMS will continue to explore variables for stratification based on availability, importance to patients and consumers, and multi-stakeholder agreement.

When providing suggestions for stratification, commenters provided additional comments, concerns, and rationale for stratifying the Overall Star Rating.

- Three commenters believed stratification would better reflect hospital care and add credibility and validity to the star ratings.
- One commenter stated that each stratum should be based on at least three measure groups that reflect the strata type.
- Two commenters stated that stratification is needed to achieve meaningful and fair provider comparisons for consumers.
- Two commenters acknowledged the added complexity of stratification to the Overall Star Rating methodology.
  - Despite added complexity, one comment believes consumers have the ability to recognize the importance of hospital classifications to make meaningful quality comparisons.
- One commenter noted other quality rating systems, such as U.S. News and World Report and Truven 100 Top Hospitals, that successfully employ stratification.
- Five commenters stated that hospitals included in the Overall Star Rating have too different numbers and types of reported measures and stratification by hospital type may provide consumers with accurate comparison information for healthcare decisions.
- Two commenters made comments specifically about academic hospitals.
  - One of those commenters noted that academic hospitals predominantly receive one-, two-, and three-star ratings.
  - One of those commenters stated that teaching hospitals provide complicated and new procedures and treatment and care for broader SDS patient population but are compared to low volume hospitals or hospitals with more homogeneous populations.
- Two commenters advocated for both retaining the Overall Star Rating with and without stratification to ensure transparency and utility for patients and consumers.
- One commenter advised using no more than three strata.
- Two commenters asked that CMS disclose the impacts of stratification and other revisions to allow stakeholders opportunity to provide constructive feedback before public reporting.

**Response:** Thank you for your comments. Please note that three measure groups are required for hospitals to receive a reliable Overall Star Rating and, unless that changes, three measure groups will still be required for each stratum should stratification be implemented. Several commenters have provided comments about the number and type of measures reported by hospitals in regard to both public reporting thresholds and stratification. Most of those comments include recommendations related to stratification and are, therefore, addressed in this section of the summary report.

Eight commenters advised against stratifying the Overall Star Rating. Some comments contained multiple points shared by others so comments are broken apart further and grouped below.

- Four of those commenters agreed with concerns that stratification may be hard to explain, confusing to consumers, and may prevent direct comparisons.
  - Two of those commenters further stated that stratification would violate the goal of the Overall Star Rating as a useful and easy to interpret tool for consumers.
  - One of those commented further pointed out that patients and consumers may not be familiar with different hospital characteristics that could be variables for stratification.
- Four of those commenters suggested addressing any measurement bias on the individual measures, rather than the Overall Star Rating.
- Instead of stratification, one commenter advocated for star ratings on the measure group level to allow consumers to select hospitals based on different aspects of quality and performance.
- One of those commenters pointed out the numerous variables for stratification and that it may be unlikely for stakeholders to come to a reasonable consensus on those variables and methods. The commenter requested that CMS conduct simulation analyses to determine any impact on each stratum of hospitals and clarify what type of methodological and display decisions will be made along with stratification.

**Response:** Thank you for your comments. The concept and potential variables for stratification or peer grouping will be brought to the Patient & Patient Family Caregiver Work Group, Technical Expert Panel, and Provider Leadership Work Group for input on how stratification would be understood or utilized by consumers as well as how to implement stratification. Many stakeholders have also expressed interest in Domain Star Ratings for each of the existing measure groups for which the investigation of feasibility is ongoing.

### 3.2. Measure Inclusion Criteria

Five commenters expressed support for the current criteria for measure inclusion.

Six commenters recommended or provided support for the removal of measures that are already excluded from the Overall Star Rating.

- One commenter supported the removal of topped out measures from the Overall Star Rating, similar to the measure policies for Inpatient Quality Reporting (IQR) and Outpatient Quality Reporting (OQR) programs.
- One commenter recommended removing overlapping measures from the Overall Star Rating, which would necessitate the removal of the PSI-90 composite measure.
- Four commenters urged CMS to exclude delayed or retired measures, especially those removed from the IQR program and for which hospitals are no longer required to collect data, from the Overall Star Rating calculation.

Three commenters provided comments and suggestions for modification of the measure inclusion criteria.

- One commenter encouraged routine assessment of the inclusion and exclusion criteria.
- One commenter recommended only including measures that have been endorsed by the National Quality Forum (NQF) to ensure the inclusion of only reliable measures for consumers' assessment of quality.
- One commenter recommended only including measures that meet the basic reliability and validity tests for public reporting.

One commenter suggested that CMS hold a public comment period before adding new measures to the Overall Star Rating.

One commenter suggested listing the predefined thresholds a hospital needs to meet for a given condition or procedure to report measures since case count impacts the reporting of measures and therefore, the inclusion of measures within the Overall Star Rating. The commenter noted that it is counterintuitive to exclude measures based on low volume considering the literature linking higher volume to better quality.

**Response:** Thank you for your comments, support, and suggestions regarding the measure inclusion criteria for the Overall Star Rating.

Please note that the current measure inclusion criteria excludes: 1) Measures suspended, retired, or delayed from public reporting on *Hospital Compare*; 2) Measures with no more than 100 hospitals reporting performance publicly; 3) Structural measures (for example, participation in a registry, or patient volume); 4) Measures for which it is unclear whether a higher or lower score is better (non-directional); 5) Measures not required for IQR Program or OQR Program; and 6) Overlapping measures (for example, measures that are identical to another measure, or measures with substantial overlap in cohort and/or outcome). In addition, the case count of each individual measure is applied prior to the inclusion of measure results within the Overall Star Rating. The measure inclusion criteria may be modified to fit the measures available on *Hospital Compare* as measures are removed and added.

The Overall Star Rating is meant to summarize and be inclusive of the existing measures publicly reported for the IQR and OQR programs on *Hospital Compare*. It is beyond the project scope to evaluate or discriminate between individual measures. The minimum case count, public reporting, NQF endorsements, and assessment of reliability and validity testing is within the purview of individual measure stewards. CMS will continue to seek feedback regarding broad criteria for measure inclusion and exclusion as the number and types of measures reported on *Hospital Compare* evolves.

## 4. Other Comments Provided

### 4.1. Inclusion of Specific Measures or Types

One commenter stated that the Overall Star Rating has an overreliance on claims-based outcome measures, which places burden on hospital resources to reconcile claims to find opportunities for improvement compared to abstracting process measures.

One commenter recommended re-considering certain structural measures for inclusion. The commenter suggested including structural measures in a measure group or as potential stratification criteria.

Two commenters expressed concern that many of the measures included in the Overall Star Rating are based on data that is from two to three years prior of public reporting. Some commenters pointed out specific concerns about including those measures.

- One commenter noted that this may be misleading to consumers since the results do not reflect current hospital performance. The commenter urged CMS to work to mitigate the lag in reported performance to better reflect real-time quality improvements.

One commenter provided comments on the inclusion of the HCAHPS survey data in the Overall Star Rating. The commenter noted research indicating a greater likelihood of low HCAHPS scores from patients admitted via the emergency department.

Thirteen commenters provided comments on the inclusion of the PSI-90 composite measure. Some comments contained multiple points shared by others so comments are broken apart further and grouped below.

- Five commenters stated that the PSI outcomes occur disproportionately in teaching hospitals or hospitals providing specialized care due to volume of surgical cases and are more reflective of high-risk procedure volume than differences in performance.
- Six commenters claimed that the PSI claims data is not clinically validated and do not represent patient-level risk factors and the quality of care at a hospital.
- Two commenters stated additional deficiencies and concerns of the PSI-90 measure, including surveillance bias, outcomes not preventable through evidence-based practices, and lack of appropriate exclusions.
- Seven commenters noted that the PSI-90 measure, with its high loading, contributes nearly all performance information to the Safety of Care measure group.
  - One commenter recommended that CMS reconsider including the PSI-90 composite and, instead, consider including the component measures so that the loading is distributed amongst individual measures.
- Five commenters recommended removing the PSI-90 measure from the Overall Star Rating.
  - Four commenters noted that the measure has been revised for the 2018 IQR program reporting and removed from HVBP for fiscal years 2019 through 2023.
    - One commenter expressed concerns that the modified version will not address concerns related to the Overall Star Rating.
- Two commenters suggested that CMS re-name and re-endorse the PSI measures, including the PSI-90 composite, to reduce consumer and provider confusion, considering the substantial specification changes.
  - One commenter noted that measures that have changed inclusion or exclusion criteria become different measures and are no longer comparable to previous quarters.

- One commenter noted the recent and specific PSI measure changes.
  - PSI-08 changed substantially from measuring postoperative hip fracture to measuring in-hospital falls with hip fracture.
  - PSI-07 (central venous catheter-related blood stream infection rate) was removed from PSI-90.
  - PSI-09 (perioperative hemorrhage or hematoma rate), PSI-10 (postoperative physiology and metabolic derangement), and PSI-11 (postoperative respiratory failure) were added to PSI-90.
  - PSI-12 (postoperative pulmonary embolism or deep vein thrombosis) and PSI-13 (postoperative sepsis rate) had specification changes.
  - PSI-15 (accidental puncture or laceration) was changed from including all medical/surgical diagnoses to abdominal-pelvic procedures only.
  - Weighting of PSI-03 (pressure ulcer rate), PSI-08, and PSI-13 increased, while others decreased.

One commenter proposed CMS strongly consider removing PSI-04 (death among surgical inpatients with serious treatable complications) from the Mortality measure group. The commenter noted the measure includes patients even if their adverse outcome was present-on-admission.

One commenter provided comments on the inclusion of the Hospital-Wide Readmission (HWR) measure. The comment recommended removing the HWR measure from the Overall Star Rating until the measure is adequately risk-adjusted for SDS status factors that are beyond the immediate control of hospitals. They further noted that the high loading of the HWR measure could provide an inaccurate Readmission measure group score and star rating.

Two commenters recommended removing the Efficient Use of Medical Imaging measure group and measures. Some commenter provided further comment about their recommendation to remove the Efficient Use of Medical Imaging measure group.

- One commenter stated that the efficiency measures don't appear to add value to the Overall Star Rating and may not be meaningful to patients.

One commenter recommended removing OP-18b/ED-3 (median time from emergency department arrival to emergency department departure for discharged emergency department patients) from the Timeliness of Care measure group given that it may inappropriately encourage speedy treat and release.

One commenter urged CMS to align the Overall Star Rating with payment programs, such as HVBP.

**Response:** Thank you for your comments and suggestions regarding specific measures and measure types. The Overall Star Rating is meant to summarize and be inclusive of the existing measures publicly reported for the IQR and OQR programs on *Hospital Compare*. Measures are included in the Overall Star Rating if they are reported for IQR or OQR on *Hospital Compare* and can be feasibly included in the star rating calculation. For example, structural measures are currently excluded from the Overall Star Rating because they do not clearly indicate good or poor performance.

Regarding specific measures of stakeholder concern, it is beyond the project scope to evaluate individual measures. The minimum case count, public reporting, endorsement, naming, and assessment of reliability and validity testing is within the purview of individual measure stewards. As measures are modified by the measures steward, the revised measure specifications are carried over within the Overall Star Rating.

## 5. Public Input Period

Twelve commenters expressed appreciation for the opportunity to comment on the Overall Star Rating during the public input period.

Two commenters requested an additional 30 days to provide comments on topics of Overall Star Rating enhancements, ongoing reevaluation, and future considerations before implementation of any refinements.

**Response:** CMS and the project developers are dedicated to public transparency and stakeholder engagement and will continue to seek expert and public input as the methodology evolves. Although CMS projects that request public input are only required to post materials for two weeks, this public input period was set for 30 days to ensure adequate time for stakeholder and public feedback. CMS is committed to ongoing reevaluation of the Overall Star Rating, which allows additional opportunity for stakeholder and public input in the future.

## 6. Overall Project

### 6.1. Project Goals and Accessibility

Eight commenters expressed their support of the project goals to provide accessible, meaningful, and usable data to consumers.

- Specifically, two commenters stated that they also believe that the project should communicate information that is reliable, valid, and useful to patients through a rigorous statistical process.
- One commenter lauded CMS' transparency to publicly disclose provider information to help promote the receipt of high-quality care by consumers and the improvement of care by providers.

### 6.2. Summarizing Information

However, four commenters expressed concern of oversimplification of a complex concept.

- Two commenters stated that combining multiple facets of quality within a single graphical representation is a reductive approach that oversimplifies the complex business of delivering healthcare and may be misleading to patients and consumers considering varied services offered at different hospitals.
- One commenters requested CMS reevaluate the appropriateness of an overall hospital star rating that oversimplifies complex and individualized choices patients must make about their care and health.
- One commenter requested CMS clarify how the Overall Star Rating differs from existing star ratings for other providers and ensure the hospital ratings do not oversimplify patient decisions while potentially exacerbating disparities in care.

Nine commenters suggested that a single rating may not be appropriate for public reporting.

- Three commenters stated that the measures on *Hospital Compare* were not meant to be summarized within a composite and displayed as a single graphical representation of hospital quality.

- One commenter noted the difficulty in achieving a single graphical representation of hospital care using a limited number of variables and statistical constructs with limitations that may not reflect overall care delivered at a hospital.
- One commenter stated that continuing to report a single star rating as representation of the many aspects of quality care is a disservice to patients, caregivers, and hospitals.
- An additional commenter added that quality can vary widely across departments within an institution and ignoring this nuance ignores the reality that a single score or ratings may not accurately represent quality for all conditions or procedures at a single hospital. The commenter added that patients are more concerned about quality relevant to the specific care that they will receive.
- One commenter said that patients' abilities to make well-informed choices are impaired by this one size-fits-all model that does not reflect the full picture of hospital care.
- Another commenter added that combining ratings across conditions and procedures obfuscates quality at the level of specificity.
- One commenter relayed that they do not believe that the Overall Star Rating reflects quality of care being delivered at their organization.

### 6.3. Other Overall Concerns

One commenter noted potential issues and unintended consequences from the Overall Star Rating. The commenter suggested low star ratings can cause burden and guilt on staff and create high hospital staff turnover and burnout.

One commenter recommended fully vetted alternative approaches to a five-star rating considering the complexity of the methodology, which is challenging to understand and less meaningful to consumers.

**Response:** CMS and the project developers thank you for your feedback on the goals and concept of the Overall Star Rating project. The goal of the project is to improve the usability and interpretability of hospital quality information by summarizing the current quality measures on *Hospital Compare* for patients and consumers. Many patient advocacy groups, government entities, and purchasers have requested clearer display options for consumers and have expressed support for the launch of the Overall Star Rating. The project developers convened and worked closely with a Patient & Patient Advocate Work Group to guide development and solicit input on the need for and the usefulness and meaningfulness of the Overall Star Rating. In addition, we also believe that the Overall Star Rating should convey reliable and useful information to consumers through a rigorous statistical process and we will continue to evaluate approaches to improve the methodology with the Technical Expert Panel and Provider Leadership Work Group.

CMS appreciates the comments regarding the complexity of creating a single star rating for the currently developed quality measures on *Hospital Compare*. Previous research has shown that consumers need help understanding hospital quality information, and prefer information to be presented in a more condensed and annotated manner. The Overall Star Rating is not meant to replace the existing measures and the public is still able to review measure-specific specifications and results for more nuanced information. Furthermore, CMS developed the Overall Star Rating methodology to complement the methodologies and goals of the other CMS programs and Star Ratings initiatives, which are reported on other CMS Compare sites. CMS will also continue to explore other mechanisms for displaying quality data.

## 7. Overall Methodology

### 7.1. Use of a Complex Methodology

Seven commenters suggested that the methods were too complex with some suggesting it could lead to further confusion. Some comments contained multiple points shared by others, so comments are broken apart further and grouped below.

- One commenter added that they were concerned about the complexity of the star ratings method, along with the unreliability of the results and inability for their members to use the ratings. They added that they would support ratings that both benefit the public and are useful to hospitals in driving their quality improvement work.
- Another commenter expressed challenges in understanding and explaining the methodology to consumers and clinicians.
- One commenter stated that the star ratings add unnecessary complexity. They added that patients and families do not possess the clinical and statistical knowledge or the time needed to decode the star ratings and extract the information that is most relevant to them.
- One commenter said if the information does not accurately account for health care quality or is not comprehensible and useful, it can lead to misinformed choices.
- One commenter added that they have heard similar uncertainties and confusion from their members about the appropriateness of the methodology and selected measures, as well as the usability for patients.
- An additional commenter voiced that they appreciate the time and work on reevaluating the star ratings, but added they remain concerned with the methodology. The commenter feels the star ratings published on the website are inaccurate and misleading to consumers that are seeking hospitals to provide their care.
- One commenter specifically mentioned how confusing the k-means clustering algorithm is and asked where the cutoff is that separates star categories.
- Three commenters specifically mentioned the highly technical nature and complexity of the latent variable modeling, making it difficult to understand and explain.
- One commenter requested resources that simplify the methodology on the QualityNet webpage.
- Another commenter wondered how many hospitals have the staff and statistical knowledge to understand the methodology and turn it into actionable improvement plans.
- One commenter stated that a simpler methodology may increase transparency and support hospitals in improving measure-level and Overall Star Rating performance.

**Response:** CMS and the project developers appreciate the comments regarding the complexity of the star ratings methodology. While the methodology may appear complex, it is important that the approach for combining multiple measures into one score is scientifically sound; that is, it is statistically meaningful and can be easily recognized and interpreted by consumers.

The k-means clustering algorithm establishes five clusters or categories of hospitals to represent the five-star ratings. The cutoffs that separate star rating categories are based on ranges of summary scores so that hospitals in one star rating category are more similar to each other and more different than hospitals in other star rating categories. The range of summary scores for each star rating category are available within the Quarterly Updates and Specifications Report on the QualityNet webpage for each quarter release of the Overall Star Rating.

CMS aims to maintain transparency. In addition to the Quarterly Updates and Specifications report, also available on the QualityNet webpage, are Frequently Asked Questions, Factsheet documents, National Provider Call materials, and the SAS pack as well as a Comprehensive Methodology Report to help stakeholders understand the Overall Star Rating methodology and results. CMS will continue to identify ways to convey the complex but necessary methodology that provides quality information to the public in an easily accessible manner. Questions about the Overall Star Rating can be sent to the Overall Star Rating inbox at [cmsstarratings@lantanagroup.com](mailto:cmsstarratings@lantanagroup.com).

## 7.2. Latent Variable Modeling

Fifteen commenters provided comments specific to the latent variable modeling methodology. Some comments contained multiple points shared by others so comments are broken apart further and grouped below.

- Four commenters noted that latent variable modeling assumes one latent trait of quality for each model or measure group and that may be an unreasonable assumption with some measure groups containing more than one latent trait of quality. This can be seen through review of the scree plots and variance explained in the Overall Star Rating methodology report.
  - Three commenters noted that there is evidence of more than one latent trait of quality with only marginal variance explained within the Efficient Use of Medical Imaging measure group.
  - Three commenters questioned whether mortality measures should be combined using latent variable modeling since hospital performance is likely not similar across divisions and conditions.
- Four commenters stated that most of the topics for enhancements or ongoing reevaluation are related to the deficiencies of the latent variable modeling methodology and questioned whether latent variable modeling is appropriate for the Overall Star Rating methodology.
- One commenter said that SAS is unable to accurately and reliably handle the large volume of data used to calculate the star ratings.
- Nine commenters commented on latent variable modeling and its use of loadings, noting that the methodology assumes a single latent variable for each measure group, resulting in negative, near-zero, and large measure loadings.
  - Two commenters noted that the HAI-6 measure has a positive but near-zero measure loading and, therefore, contributes very little to the Safety of Care measure group.
    - One commenter pointed out that the HAI-6 measure likely reflects a different, but equally important, aspect of quality than that assigned to the measure group, resulting in little contribution of that measure to the measure group and Overall Star Rating.
  - One commenter focused on the Efficient Use of Medical Imaging measure group, which has two measures with small negative loadings and three other measures, with OP-10 (abdomen computed tomography use of contrast materials) contributing the most with the highest measure loading, that make up two thirds of the total measure group score.
  - Two commenters noted that the measure loadings change each quarter and it is difficult for hospitals to determine where to focus their improvement efforts.
  - One commenter noted that small or negative loadings is indicative of a fundamental problem with the way the latent variable model summarizes the latent trait of quality. Measures with low or negative loadings do not reflect the latent trait.

- One commenter suggested adding additional latent variables or measure groups to effectively reflect the number of latent traits based on the included measures.
- Eleven commenters advocated for alternative approaches to latent variable modeling.
  - Six commenters recommended removing latent variable modeling from the methodology and, instead, using consistent weights for each measure that are evaluated each year. One commenter stated that this may provide scoring stability, transparency, and results that are easier to interpret for hospitals and consumers.

**Response:** CMS appreciates these comments as well as several of the proposed ideas for future evaluation. The original selection of latent variable modeling (LVM) was based on the input of the original Star Ratings Technical Expert Panel as well as additional statistical consultants. The use of LVM offered several advantages for summarizing individual measures including accounting for consistency of performance by giving more importance to measures that are correlated within a group, accounting for missing measures by using all available information to generate group scores, and accounting for sampling variance as a result of differences in hospitals' volumes.

CMS acknowledges that the strength of the Efficient Use of Medical Imaging LVM is limited by the absence of a strong latent trait in comparison to other models. CMS sought input on this in earlier public input periods and included this measure group given its importance to patients and consumers, albeit at the lowest 4% weight given known limitations. CMS will consider measure groupings and measure group inclusion as part of regular reevaluation activities. In the case of the mortality measure group, earlier analyses demonstrate a dominant latent variable common to all mortality measures included within the group. While each LVM may not capture all elements of quality measured by each individual measure, the models seek to capture common information that reflects quality for a given measure group. In other words, hospitals performing well on individual measures perform better within the measure group while hospitals that perform poorly on individual measures perform poorly within the measure group. Several validation and testing analyses support this approach.

Regarding the low loadings of select measures in select groups, several explanations are likely, including smaller denominators for such measures, less common association with other measures in the measure group, and more limited reporting profiles. The LVM approach accommodates this degree of missing information or less reliable information by applying a lower loading to such measures based on the information available in the data. The use of individual measure weights as suggested by one comment would likely not be feasible given challenges in establishing consensus and regularly maintaining such individual measure weights across a multitude of measures and measure groups. Regarding the effect of coding behavior on measure performance, CMS seeks to include valid and reliable measures based on broad inclusion and exclusion criteria. In addition to regular audit procedures conducted by CMS, each included measure has been previously vetted by NQF to ensure validity. Ultimately, CMS believes that use of the LVM approach creates a consistent approach to summarizing hospital quality information within each measure group while accommodating several policy decisions in assigning star ratings.

### 7.3. Clustering into Star Categories

One commenter expressed concerns with fitting a normally distributed, continuous rating from seven latent variable models based on a weighting scheme into five clusters.

Four commenters provided comments on the curve-based method of k-means clustering. In other words, k-means clustering identifies five clusters of hospitals so there will always be one- and two-star hospitals regardless of hospital performance improvements.

- One commenter cited a situation in which value-based market is based on financial rewards given only to the highest performers, and not those that improved quality. Rewards based on a curve, rather than meaningful thresholds, will prevent high-quality hospitals from being rewarded and could discourage hospitals from sharing best practices.
- One commenter supposed a hypothetical situation in which performance uniformly declines across all hospitals. In this scenario, the current star ratings would remain relatively the same when, in actuality, all hospitals should be receiving a low star rating.
- One commenter suggested a threshold-based methodology, rather than k-means clustering with or without stratification of hospitals.
- One commenter said the current methodology requires a certain percentage of hospitals in each of the star categories. They added that hospitals may receive a star rating in a certain category even though the quality of care provided may not be meaningfully different from that provided from hospitals in a higher category.

One commenter expressed concern about the use of confidence intervals to determine star ratings. Smaller hospitals are more likely to have longer confidence intervals, resulting in a better star rating, compared to larger hospitals that have shorter confidence intervals.

**Response:** CMS appreciates your comments and concerns. Please note that the Overall Star Rating is not used for hospital payment determination and current methodology does not utilize confidence intervals for assignment of star ratings but does use confidence intervals for the assignment of performance categories for each measure group to mirror the approach used to determine above or below average outliers for individual measures.

CMS considered several approaches for translating summary scores to star ratings. K-means clustering is a standard method for creating categories (or clusters) so that the observations (or scores) in each category are closer to their category mean than any other category mean. There are no designated percentages of hospitals for the star categories. The goal of clustering is to ensure that the hospital summary scores in each category are “more like” that of the other hospitals in the same category and “less like” the summary scores of hospitals in other categories. The approach to star rating assignment is a comparative approach but accommodates shifts in the distribution. CMS continues to evaluate the stability and impact of the clustering methodology and will potentially explore alternative options with the Technical Expert Panel in the future.

#### 7.4. Measure Groups

One commenter suggested renaming the Mortality measure group to “Survival,” considering that “above the national average” can result in confusion and be seen as a greater number of deaths versus higher performance.

One commenter stated that the separation of measures into measure groups for the Overall Star Rating may be unnecessary and awkward. The measure groups and weighting of those measure groups as well as the latent variable modeling make assumptions about which measures are better indicators of quality when there is not a correct answer.

**Response:** Thank you for your comments. The measure groups names were assigned with the help of our Patient & Patient Advocate Work Group. While some work group members originally agreed that survival makes more sense considering the directionality of the measures, other work group member felt that a “Survival” measure group could be misleading. The project developer will bring this suggestion back to the work group and CMS for consideration.

### 7.5. Domain Star Ratings

One commenter stated that the current approach of an Overall Star Rating does not help consumers make healthcare decisions for specific conditions or procedures.

Fourteen commenters advocated for Domain Star Ratings or star ratings for subsets of measures on *Hospital Compare*.

- One commenter stated that Domain Star Ratings would temper conflicting priorities among stakeholders, increase transparency, and increase value to consumers.
- One commenter stated that Domain Star Ratings would be more meaningful and actionable for both consumers and providers.
- One commenter stated that the consumer should be able to make decisions based on which aspects of care are most significant for their specific situation, rather than a single composite.
- One commenter stated that the measures on *Hospital Compare* were never meant to be combined to create a single score of hospital quality and, in some cases, the Overall Star Rating does not reflect aspects of care relevant to patient needs.
- One commenter pointed out that Domain Star Ratings would allow consumers to select hospitals based on different aspects of performance, rather than just location, hospital type, and emergency services.
- One commenter noted that Domain Star Ratings, accompanied by specific performance thresholds, may allow the same measures to be used when making comparisons across hospitals.
- One commenter requested displaying Domain Star Ratings prominently alongside the Overall Star Rating.
- One commenter requested that Domain Star Ratings replace the Overall Star Rating on *Hospital Compare*.
- Three commenters requested Domain Star Ratings be based on subsets of measures of importance on *Hospital Compare*, such as conditions, procedures, inpatient, outpatient, emergency department, patient safety, patient experience, and cardiac care.
- Three commenters requested that CMS provide star ratings for each measure group, similar to the HCAHPS star rating.

**Response:** CMS and the project developers appreciate the feedback on more specific areas of focus for star ratings. CMS will take these comments into considerations in any future projects.

### 7.6. Customizable Star Ratings

One commenter suggested that consumers should be able to filter hospitals on *Hospital Compare*, based on the different aspects of performance through measure groups and other features beyond location, hospital type, and emergency services.

Another commenter advocated for consumer-driven weights, in which the consumer determines which measures hold the most importance based on their individual preferences, values, and needs.

**Response:** CMS and the project developers appreciate the feedback on more customizable views of *Hospital Compare*. CMS will take these comments into considerations when evaluating future options for the *Hospital Compare* website.

### 7.7. Transparency & Stakeholder Engagement

One commenter expressed appreciation of CMS' efforts to encourage transparency in care delivery across the entire health care industry and support sharing meaningful quality information with patients and consumers.

Eleven commenters provided suggestions on how CMS can communicate improvements to star ratings, as well as recommendations on how to guide hospital improvement.

- One commenter added that the Overall Star Rating does not provide actionable information for hospitals to identify opportunities for improvement.
- One commenter recommended that CMS continue ongoing investigation for areas of future improvement for the Overall Star Rating and convene stakeholders regularly.
- One commenter stated that any additional changes to the Overall Star Rating methodology should ensure that consumers can better understand differences among hospitals.
- One commenter requested that CMS hold several educational webinars on the proposed changes in the methodology before implementation.
- One commenter stated that the current approach of an Overall Star Rating does not promote meaningful quality improvement across hospitals.
- Three commenters urged CMS to release the research database on the proposed methodology enhancements to allow stakeholders the opportunity to review the data, understand the changes and any impact on results, and communicate any concerns or recommendations.
- One commenter noted the competing number of five-star rating systems, especially in California, which include varying methods and measurement periods. Competing ratings divert hospital resources for quality improvement and confuse consumers.
- One commenter noted the importance of any changes to the display of data by star categories to accurately reflect the quality of care provided by hospitals to their patients.
- One commenter noted broad measures, such as Mortality and Readmission, can be misinterpreted by patients and more useful measures may exist. Context should be provided to patients regarding the score relative to other hospitals when broad measures are used that are influenced by intensity of the illness or hospital specialty.

Nine commenters expressed concerns regarding the reevaluation process for Overall Star Ratings.

- Four commenters expressed concern that the reevaluation of the Overall Star Rating is being conducted by the same contractor that developed the Overall Star Rating.
- One commenter advocated for CMS to consider other contractors for evaluation of the methodology and increased external stakeholder involvement during the evaluation of the Overall Star Rating.
- Four commenters requested that an independent third-party review the methodology to ensure meaningful and accurate star ratings that do not disproportionately disadvantage any hospital types.

One commenter requested more complete impact analyses and statistical model outputs for each proposed enhancement. Specifically, the commenter requested:

- The number of hospitals that change star rating and whether it is an increase or decrease in star rating;
- How the enhancements impact the measure loadings;
- The change in hospital summary score cutoffs for each star rating category;
- The number of hospitals whose summary scores were previously winsorized;
- The change in explanatory power of the model (for example, R-square); and
- The influence of outliers on the clustering of hospitals.

Four commenters requested that the Overall Star Ratings continue to be suppressed while CMS further reviews the methodological enhancements through additional testing and stakeholder engagement, with one commenter adding this could help ensure a meaningful and accurate assessment of quality without further confusion.

Two commenters advocated for greater alignment between the Overall Star Rating and other CMS programs.

Four commenters requested a more robust stakeholder engagement process.

- One commenter specifically requested close stakeholder engagement to ensure an accurate, statistically sound methodology and prevent the misclassification of hospitals to star rating categories.
- One commenter specifically suggested the CMS engage with academic medical centers and hospital associations to refine the methodology.
- One commenter suggested that CMS hold focus groups with hospitals, physicians, patients, families, and caregivers to understand how well the statistical information and star rating display are conveyed, understood, and used by all stakeholders.
- One commenter suggested that CMS further test the methodology and convene panels of stakeholders to comment on proposed methodological changes.

Two commenters encouraged improvement in the data for star ratings, with one commenter recommending alignment with data periods and another advocating for more timely data.

One commenter stated that healthcare information should be readily available to patients for the purposes of improving quality in health care, expanding consumer engagement in health care decision-making, and to improve federal program administration of health care benefits.

**Response:** Thank you for your comments. CMS is dedicated to transparency and stakeholder involvement. Efforts to increase transparency and stakeholder engagement for the Overall Star Rating included, convening two separate Technical Expert Panels, a Patient & Patient Advocate Work Group, and a Provider Leadership Work Group, holding three public input periods, hosting two National Provider Calls, and allowing stakeholders to review the methodology and hospital-specific results through a national dry run. In addition, CMS has made public each release quarter the Overall Star Rating SAS pack and input file to allow stakeholder replication of results. CMS and the project developers will continue to engage stakeholders on methodological and policy decisions through panels, work groups, and public input periods.

## 7.8. Hospital Characteristics

One commenter expressed concerns that star ratings places an unwarranted risk where large hospitals, teaching hospitals, and those hospitals that serve a high proportion of low-income patients will receive lower star ratings, while still providing quality care, often to the most vulnerable populations.

Another commenter added that certain hospital types may receive lower star ratings because of these factors that are unrelated to the quality of care they provide.

Another commenter believed there is the distinct risk that larger hospitals, teaching hospitals, and hospitals serving low-income patients receive lower star ratings despite providing quality care, often to disadvantaged populations.

**Response:** CMS will continue to monitor and reevaluate the star ratings methodology, including monitoring of hospital distribution in star ratings. As aforementioned, CMS will also consider these comments in future discussions regarding hospital stratification, or peer group comparison.

## 8. Other Comments (Beyond Scope)

### 8.1. Beyond Scope Comments

Eleven commenters suggested that CMS consider risk adjustment for SES factors in the measures included in star ratings or the star ratings methodology.

**Response:** The current star ratings project does not update or change the existing measure methodology. NQF has held several measure-level evaluations to determine whether SES adjustment should be incorporated into measure specifications, resulting in the maintenance of several outcomes measures without SES risk adjustment. However, if SES risk adjustment is included in the measure-level methodology, this will be reflected in the scores that are included in star ratings.

One commenter added that CMS finalized a risk adjustment methodology for the Hospital Readmission Reduction Program (HRRP) that allowed separate comparisons of hospitals based on a facility's proportion of dual-eligible patients. They added that stratification is not risk adjustment and recommended more be done to account for these factors.

**Response:** The comment about HRRP is currently beyond scope for this project. Exploration of this variable would be at the measure level. Any changes to the measure-level results would be reflected in the scores included in star ratings.

Some of the comments received are considered beyond scope as they did not address star ratings.

- One commenter noted that hospital surveys by State and Federal Regulators can be a mechanism for correction.
- Four commenters provided comments on the current system of quality measurement.
  - One commenter supported CMS' effort to incentivize quality through outcome-based measurement, but believes current measurement should be revised to account for medical complexities of patients and SES challenges that hospitals face in caring for patients.
  - One commenter advocated for measurement that increases quality and efficiency with an emphasis on wellness and prevention, care coordination, and the use of evidence-based medicine, while utilizing consumer choice and competition to enhance value.

- One commenter supported the alignment and streamlining of meaningful measures across healthcare programs and suggested that CMS focus on measures that matter most for improving patient care and outcomes, without distracting limited resources from ongoing improvement, patient care, and innovation.
- One hospital noted *Hospital Compare* may not reflect the care of greatest interest to patients, for instance outcomes for emergent surgical care or cancer care follow-up or treatment may be more important than other measures currently available.
- One commenter requested that the HVBP program reduce weights for the PSI-90 and Hip/Knee Complications measures due to the delay between data collection and reporting.
- Three commenters provided comments about measure-level methodologies.
  - One commenter stated that the predicted over expected approach complicates hospital and consumer understanding of how the observed values evaluate hospital performance. Hospitals may influence measure scores by improving the documentation and coding of comorbidities that count towards the predicted and expected value calculations.
  - One commenter recommended that reliability-adjusted outcome measures exclude low volume hospitals whose results appear average due to the wide confidence intervals, resulting in unreliable results with inaccurate performance measurement. The commenter believed that these hospitals results may distort the measure group scores and star ratings of hospitals with adequate case counts for reliable results. Therefore, the commenter requested that CMS replace the public reporting thresholds with a requirement of minimum hospital-specific blended ratio.
  - One commenter urged CMS to:
    - Assess HCAHPS' effectiveness in measuring hospital performance on patient satisfaction based on what patients care most about, how hospital characteristics impact HCAHPS scores, and regional trends in satisfaction;
    - Modernize and improve the survey by providing notice to patients, providing online and mobile surveys, and administering the survey on-site; and
    - Adjust survey mode for potential impact of how hospitals are surveying patients.
- Four commenters recommended that CMS properly risk-adjust outcome measures and account for SES.
  - One commenter requested CMS risk adjust the mortality, readmission, and hip/knee complications measures for dual Medicare and Medicaid eligibility as an indicator for low SES. The commenter stated that this may reduce the need for stratification of the Overall Star Rating since hospital scores will be more comparable on the measure level.
  - One commenter requested that CMS partner with experts in the private sector and adjust readmission measures, in particular. Without adequate measurement and adjustment, efforts to reward higher performing hospitals may result in lower funding for those that serve more vulnerable populations.
  - One commenter expressed concern about 30-day readmission measures and urged for their suspension or retirement due to the influence of social determinants of health. The commenter suggested incorporation of SES/SDS factor risk adjustment and hospital stratification and recommended consideration of a 7-day post-discharge window, instead of the current 30-day window.
  - One commenter requested that CMS properly risk adjust outcome measures so that hospitals who care for more sick and complex patients are not unfairly rated within the Overall Star Rating.

- One commenter conducted analyses on the HCAHPS survey data and found that certain type of hospitals, such as large hospitals, teaching hospitals, and hospital serving a high proportion of low income patients, were more likely to receive a lower HCAHPS star rating. The commenter also found a greater likelihood of low HCAHPS scores from patients admitted through the emergency department so hospitals with higher emergency department volumes might score lower even though their quality may be better than hospitals with lower emergency department volumes.
- Two commenters expressed concern of using older data that may not be reflective of current performance or initiatives.
  - One commenter recommended a yearly time-period.
  - One commenter noted the scores are misleading to consumers because they do not reflect current hospital performance and urged CMS to mitigate the lag time to reflect current quality improvement efforts.

**Response:** Measure-level development, risk adjustment, case count requirements, and data collection timeframes are within the purview of CMS and each individual measure methodology and steward. The purpose of the Overall Star Rating is to summarize the existing measures on *Hospital Compare* and it is beyond the project scope to evaluate or dictate changes to individual measures reported on the *Hospital Compare* webpage.

### **Overall Analysis of the Comments and Recommendations**

CMS will continue to evaluate the current Overall Hospital Quality Star Ratings. The star ratings are designed to be a complementary display to the existing publicly reported measures on *Hospital Compare*. As measures and display options evolve and input from stakeholders is received, CMS will continue to consider improvements to the Overall Star Ratings methodology. CMS and the project developers will continue to hold public input periods, Technical Expert Panels, and Work Group meetings to ensure broad stakeholder engagement. Other questions or comments can be sent to the following email address at any time: [cmsstarratings@lantanagroup.com](mailto:cmsstarratings@lantanagroup.com). CMS is committed to evolving the star ratings methodology and display through reevaluation work and public input through transparent channels.

**Appendix A. Measure Group Weighting Tables**

*Table 1. Current Measure Group Weighting for the Star Rating*

<b>Measure Group</b>	<b>Star Rating Weights</b>
Mortality	22%
Safety of Care	22%
Readmission	22%
Patient Experience	22%
Effectiveness of Care	4%
Timeliness of Care	4%
Efficient Use of Medical Imaging	4%

*Table 2. Example of Alternative Measure Group Weighting*

<b>Measure Group</b>	<b>Star Rating Weights Alternative #1</b>	<b>Star Rating Weights Alternative #2</b>
Mortality	20%	25%
Safety of Care	20%	15%
Readmission	20%	15%
Patient Experience	20%	25%
Effectiveness of Care	8%	8%
Timeliness of Care	8%	8%
Efficient Use of Medical Imaging	4%	4%