

PUBLIC COMMENT SUMMARY REPORT

Project Title: Overall Hospital Quality Star Rating on *Hospital Compare*

Dates:

The Call for Public Comment ran from February 28, 2019 to March 29, 2019. The Public Comment Summary was made on June 24, 2019.

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) had initially contracted with Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE) and Lantana, Inc. to reevaluate the Overall Hospital Quality Star Rating on *Hospital Compare* under Development, Reevaluation, and Implementation of Outcome/Efficiency Measures for Hospital and Eligible Clinicians, Option Period 5; contract number: HHSM-500-2013-13018I, Task Order HHSM-500-T0001 and MIDS III *Hospital Compare* Support Contract (HCSC); contract number: 75FCMC18D0029/75FCMC18F0001. CMS is currently contracting with Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE) and Lantana, Inc to continue the reevaluation of the Overall Hospital Quality Star Ratings on *Hospital Compare*. The contract name is Development, Reevaluation, and Implementation of Outcome/Efficiency Measures for Hospital and Eligible Clinicians, Base Year; contract number HHSM-75FCMC18D0042, Task Order Number HHSM-75FCMC19F0001, and the Lantana contract is called MIDS III *Hospital Compare* Support Contract (HCSC); contract number HHSM-500-2013-13010I/HHSM-500-T0001. As part of the reevaluation process, CORE had requested interested parties to submit comments on potential methodology updates and areas of future work for the Star Rating.

Project Objectives:

The primary goal of this project is to reevaluate the methodology for the Overall Hospital Quality Star Ratings to improve the usability and interpretability of *Hospital Compare* for patients and consumers.

Star Ratings provides patients and consumers with a single tool to inform them about multiple dimensions of quality, represented by the existing measures on *Hospital Compare*, and capable of incorporating new measures that may be added in the future.

The purpose of this request for public comment was for CMS to gain feedback from a broad range of stakeholders (including technical experts, providers, patients, purchasers, and the public at large) on an array of potential updates to the Star Rating methodology, described in detail below.

Information About the Comments Received:

Public comments were 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>).

One hundred and forty-five responses were received from 140 commenters during the public input period. Specifically, we received comments from:

- 24 Hospitals
- 40 Individuals
- 35 Health systems
- 16 Hospital associations
- 3 Professional associations
- 15 Medical universities
- 6 Healthcare performance improvement organizations
- 1 Purchaser

Stakeholder Comments—General and Project-Specific

CMS sought comments on nine specific potential changes to the Overall Hospital Quality Star Ratings methodology, and comments are organized to show responses to those items. However, many commenters also had statements or questions about other aspects of the Overall Star Ratings as well as other CMS programs.

- Public Input Period Topics
 - Public Comment Process
 - February 2019 Methodology Updates
 - Measure Grouping
 - Incorporating Measure Precision
 - Period-to-Period Shifts
 - Peer Grouping
 - Closed-Form Solution
 - Long Term Future Considerations
 - Explicit Approach
 - Alternatives to Clustering
 - Incorporation of Improvement
 - User-Customized Star Rating
- Additional Comments about Star Ratings
 - Overall Project
 - Overall Methodology
 - Latent Variable Modeling
 - Stakeholder Engagement
 - Measures Included in Star Ratings
 - Measure Group Weights
 - Incorporating Socioeconomic Risk Adjustment
 - Alignment with Other Quality Metrics
 - Display of Star Ratings
- Comments Beyond the Scope of Star Ratings

Public Input Period Topics

Public Comment Process

Fifty-eight commenters appreciated the opportunity to provide comments on the Overall Hospital Quality Star Ratings.

- One commenter complimented CMS and CORE for a frank, clear, and detailed set of materials of alternatives and improvements to the methodology.
- While they appreciated the opportunity to comment, one commenter stated that the materials were technical, which made it challenging for stakeholders to interpret.
- Another commenter congratulated CMS and CORE for seeking input to make *Hospital Compare* more useful and representative.
- One commenter appreciated CMS's dedication to improving the Star Rating.
- One commenter requested CMS continue to use the public input process to analyze the impact of the methodology on different types of hospitals and provide transparent information regarding the methodology.
- One commenter stated appreciation for CMS's ongoing receptiveness to feedback on the methodology and the ongoing engagement between CMS, industry stakeholders, and subject matter experts. The commenter added that continued conversations regarding accuracy and utility of the methodology will promote the best use of the Star Ratings and *Hospital Compare*.
- Two commenters appreciated CMS's efforts to revise the Star Ratings for more accurate results, more appropriate comparisons of hospitals, and provide ratings that have value to consumers. The commenters further elaborated that their health system could use the improved methodology to inform efficient allocation of finite resources to meet health and safety needs of their community, which includes low-income and uninsured patients.
- One commenter expressed confidence that CMS will achieve more equitable Star Ratings that fit the need of patients and families.
- Three commenters encouraged CMS to review the feedback received and make appropriate modifications to the methodology.
- One commenter further appreciated CMS's willingness to act when it is clear improvements can be made with the ratings.
- One commenter supported CMS's efforts to improve the Star Ratings as well as many of the proposed methodology updates.
- Another commenter appreciated the transparency regarding the current limitations of the methodology.
- One commenter noted they hope the comments are useful to CMS to make Star Ratings more useful to rural and CAHs (CAHs).
- One commenter appreciated the time and effort spent to address the issues outlined in the materials.
- One commenter stated they were pleased CMS was considering methodology changes to ensure stability and considering adjustments for factors outside of a hospital's control.

Two commenters provided general feedback on the proposed methodology updates included within the public comment materials.

- One commenter further noted that the topics outlined in the materials prove that CMS understands that the ratings are unstable and unreliable. They hope that the proposed methodology updates will improve the transparency, consistency, and uniform understanding of the Star Ratings.
- One commenter appreciated the proposed methodology revisions, however noted difficulty weighing in on the options without considering how they may impact one another or lead to unintended interactions.

Twenty-one commenters provided other feedback regarding the public comment period.

- One commenter recommended their comments on principles for Star Ratings be used for the development and revisions of any CMS star rating program.
- Six commenters suggested that CMS evaluate public comments from other commenters, including the American Hospital Association, America's Essential Hospitals, Texas Health Systems, New Jersey Hospital Association, and Association of American Medical Colleges.
- Thirteen commenters encouraged CMS to contact them with questions and comments or offered their services to help improve Star Ratings.
 - One commenter specifically offered their services as a representative for CAHs.
 - Another commenter specifically offered to help develop methodology alternatives for CAHs.
 - One commenter specifically mentioned that their staff are experts in comparing hospital performance within their organization.
- One commenter stated the public comment materials do not mention 'physician', 'clinician' or 'hospitals'.

Response: Thank you for your comments, feedback, and support. CMS is dedicated to improving the Star Ratings as well as engaging and being responsive to stakeholder input. CMS and CORE will review and consider your comments in conjunction with input provided from the TEP and work groups.

The current methodology is based on a series of research, analytic work, policy considerations, and numerous stakeholder engagement activities. However, CMS acknowledges that there are other methods worth exploring that may address some of the limitations of the current methodology.

As acknowledged within the public comment materials, the Star Rating methodology was not updated but measures were refreshed and updated for the July 2018 confidential preview period, resulting in unexpected shifts in star ratings. Analyses show that the Star Rating methodology is sensitive to changes in the underlying measures.

The topics outlined within the public comment materials were geared towards concerns previously expressed by stakeholders, both since the development of the Star Ratings, such as comparing like to like hospitals, as well as in response to the sensitivity of the methodology to changes in the underlying measures observed during the July 2018 confidential preview period. These topics were also vetted with the Star Ratings Technical Expert Panel (TEP), Provider Leadership Work Group, and Patient & Advocate Work Group prior to public comment.

We acknowledge that some of the topics covered within the public comment materials are very technical. We do not expect all stakeholders to be able to provide comments on all topics, such as the

closed-form solution. CMS is dedicated to transparency and engaging stakeholders in the reevaluation process. The more technical topics were first vetted through the TEP convened by CORE and the current public input period did result in meaningful comments from statistical experts.

Input received in response to the public comment period will be reviewed and considered. As in the past, CMS seeks to provide hospitals with confidential information, as well as advance opportunities to review and comment on any methodology changes, prior to national reporting.

February 2019 Methodology Updates

Three commenters were disappointed with CMS's decision to report star ratings in February 2019.

- Two commenters stated that CMS reported Star Ratings in February 2019 without any of the methodology updates requested by stakeholders through the 2017 public comment period.
 - One commenter added that the methodology continues to be complex, making replication of results difficult. The commenter instead felt CMS should have removed Star Ratings from *Hospital Compare* while acknowledging updates were underway.
- One commenter noted that CMS is concurrently soliciting feedback on the methodology, showing that the methodology requires revision.

Two commenters noted concerns with the latent variable model (LVM) used in the February 2019 release.

- One commenter stated that the February 2019 release demonstrated lack of validity for LVM. The commenter stated that negative measure loadings indicate that there is a lack of a single latent factor, and should prompt consideration of a model that does not create tradeoffs.
- Another commenter stated the two methodology updates made in February 2019 did not address stakeholder concerns regarding the LVM model, and may have introduced more instability. The commenter stated that the choice to use LVM indicates modeling selection bias, and produces unreliable measure loadings and Star Rating results.
 - The commenter added that an in-depth analysis identified model fit performance opportunities across four of the seven measure groups using the current 1-factor LVM approach. The commenter assessed the approach with four common model fit statistics and noted results based on their analyses. The commenter recommended instead examining 1-factor reduced approaches or 2-factor modeling, but acknowledged the added complexity this would bring to the methodology.
 - The commenter stated that although none of the measures were removed because of statistically significant negative loadings in February 2019, one measure had a non-significant negative loading that can still penalize hospitals for performing well on a measure.
 - The commenter also noted that while the change in the Hospital Acquired Infection (HAI) denominators led to measure loadings more similar to previous releases, unlike the measure loadings in July 2018, even with simulation analyses, it is unclear what caused the large shifts in Star Ratings from July 2018 to February 2019. The commenter speculated the shift in Star Ratings is likely because of the LVM.

One commenter recommended a deep dive into HAI data and volume-based denominators.

Response: Thank you for your comments. Since the observed shifts in Star Ratings during the July 2018 confidential preview period, CMS hosted nine listening sessions to solicit stakeholder suggestions. CMS then identified several potential methodology updates based on feedback from the listening sessions and CORE presented those topics to the TEP and work groups prior to the February 2019 release of Star Ratings. In general, the TEP and work groups were supportive of the removal of measures with statistically significant negative loadings and use of alternative HAI denominators. The TEP and work groups expressed interest but had mixed or uncertain input on other methodology updates. The methodology enhancements applied in February 2019 did result in more model stability as evidenced by loadings more consistent with prior periods. CMS will continue to reevaluate the methodology, including alternatives to the LVM as outlined in these public comment materials, to ensure that future refinements fulfill the principles of the star ratings methodology and address stakeholder concerns.

Measure Grouping

Twenty-nine commenters supported the proposed three-step approach to evaluating measure groups, including the criteria of 1) initial clinical grouping; 2) confirmatory factor analysis; and 3) ongoing active monitoring.

- One commenter stated they appreciated the revised criteria, three-step approach, and evaluation of loading balance and consistency. The commenter added that the proposed updates would mitigate concerns raised about imbalance among loadings within measure groups. The commenter advocated for a simpler methodology, more transparent methods for calculating Star Ratings, and felt these updates represent a substantive improvement.
- Two commenters urged CMS to undertake further analysis on how to improve measure groupings.
- One commenter supported clinical groupings, confirmatory factor analyses, and ongoing active monitoring since a valid use of LVM calculations require proof of correlations between measures, which should be presented to the public.
- One commenter felt the three-step approach was reasonable but not suitable. They recommended using quantitative criteria to determine and evaluate measure groups and requested further guidance on how balance and consistency would be determined in each group.
- One commenter supported all aspects of the 3-step approach except for using the balance and consistency of measure loading coefficients, as selecting measure groupings based on statistical criteria is likely to misalign with clinical groupings, and inconsistent loadings may persist regardless of regrouping.
- Three commenters provided specific comments about clinical grouping (criterion 1).
 - Two commenters agreed that measure groups should be based on clinical coherence.
 - One commenter noted measure groupings should also be relevant to the consumer.
- Eight commenters provided specific comments about confirmatory factor analyses (criterion 2).
 - One commenter suggested confirming the stability of each LVM model using bootstrap analyses. The commenter added that the test should include confirmation analyses as well as conventional indicators of model fit, such as comparative fit index, the Tucker Lewis Index, or root-mean-square error of association and stability tests.

- One commenter asked CMS to modify all measure groups where more than one dominant factor exists.
- One commenter supported the empiric measure grouping criteria but suggested not requiring a single underlying factor since it is unnecessary, not currently reflected within Safety of Care, and could cease to be reflected within other measure groups as a result of slight changes in measures. In addition, assuming one underlying factor causes issues with measure loadings.
- One commenter said that the confirmatory factor analysis is important and should ensure one dominate factor per measure group.
- Two commenters added that confirmatory analyses test the assumptions of the LVM, but also provide a stronger empirical basis for the measure groups. One commenter said that the empirical criteria for measure groups is one of the proposed changes that appears to partially address stakeholder concerns.
- One commenter stated scree plots should be consistent or similar across all measure groups.
- One commenter stated factor weights and the overly-sensitive LVM needs to be re-assessed with every rating release, and the LVM does not fit the underlying data.
- Seven commenters provided specific comments about ongoing active monitoring (criterion 3):
 - One commenter felt that “ongoing monitoring” needs to be better defined. They added an annual refresh will require a balance between predictability and flexibility for measure adaptability.
 - One commenter supported using loading consistency to evaluate measure groupings, unless there are substantial measure changes that cause changes in loadings.
 - One commenter added that without balance and consistency, confidence in the Star Ratings will wane amongst the public.
 - Two commenters supported actively monitoring measure groupings for consistency in how much each measure influences the measure group score over time.
 - One commenter also added that loadings are the most difficult part of Star Ratings to explain and CMS should use balance and consistency for loadings as a factor in evaluating groups.

Thirty commenters supported measure regrouping.

- One commenter added that regrouping may address issues within the Safety of Care group.
- One commenter noted groups should be reassessed as measures change on *Hospital Compare*.
- One commenter suggested grouping measures by obstetrical outcomes, surgical outcomes, and infections.
- One commenter supported measure regrouping and any methodology updates that would produce balanced and predictable loadings between periods.
- One commenter advocated for fair ratings that do not reflect or combine unrelated or unknown metrics.
- In a survey of a Patient Family Advisory council, one commenter noted participants support measure regrouping, despite that continuous regrouping could make it difficult for hospitals to understand their scores over time. They support the current clinical groupings, including the

Patient Experience group. Fifty percent of participants felt metric weighting was unnecessary and 75% felt that the underlying latent traits were important to publish on *Hospital Compare*. Their hospital staff supported refining the clinical coherence of the Safety of Care measure group.

- One commenter suggested the following group changes:
 - Adding VTE-6 and SEP-1 to the Safety of Care measure group;
 - Combining Effectiveness of Care, Timeliness of Care, and Efficient Use of Medical Imaging into one group with a weight of 10%;
 - Increasing Mortality and Patient Experience measure group weights to 25%;
 - Decreasing the Safety of Care and Readmission measure group weights to 20%; and
 - All metrics should have an equal weighting within a measure group.
- Eleven commenters supporting regrouping measures within the Safety of Care group.
 - One commenter urged CMS to consider removing the Patient Safety and Adverse Events Composite (PSI-90) from the Safety of Care measure group.
 - One commenter was hopeful that regrouping would minimize the measure loading of PSI-90, which they believe is misleading. However, they felt separating surgical from medical safety of care measures makes sense to providers, but not to consumers.
 - Two commenters said that the current Safety of Care measure group and the proposed groupings did not provide an adequate solution. These commenters added CMS should consider removing the PSI-90 and Hip/Knee Complication measures.
 - Two commenters suggested a broader distribution of measure groups.
 - One commenter added that if the LVM is kept, testing would be needed to see if added or removed measures changed the distribution of measure loadings. The commenter used Safety of Care as an example, adding that the six HAI measures could be moved to a group called “Safety-Infections” while PSI-90 and Hip/Knee Complications could be grouped as “Safety-Surgical.”
 - One commenter supported regrouping but felt the alternative recommendations for the Safety of Care group to be ineffective because the groupings do not meet the criteria set forth for evaluating measure groups within public comment materials.
 - One commenter supported regrouping but recommended that CMS conduct further analyses before implementing changes. The commenter did not support the proposed alternative for the Safety of Care measure group or the use of the PSI components instead of the PSI-90 composite.
 - One commenter added that the current grouping is more suitable than the proposed alternatives stating that PSI-90 should be used consistently across programs.
 - One commenter did not agree with the current loading structure, stating it is difficult to understand or explain. The commenter added that the Safety of Care group has many components that do not have bearing on the Star Ratings and the loadings for this measure group needs to be more balanced.
- Eleven commenters felt the Safety of Care group should be split in two measure groups.
 - One commenter suggested including other measures to make the measure group more statistically robust.

- One commenter noted the Surgical Safety group may be misleading to consumers by conveying that it covers all surgeries.
- Three commenters suggested using the PSI component measures.
- Five commenters supported splitting PSI-90 components into a medical and surgical group.
- One commenter felt that the regrouping options do not solve the issues, but requested CMS explore using each PSI component.

Three commenters did not support any of the regrouping options.

- One commenter said to not change the measure groups as they make sense to consumers. They added a request to change the modeling and weight approach for the Star Ratings. The commenter felt the issue was not in the grouping, but that using a LVM depends on correlation. The commenter added that CMS needs a new approach that is not based on correlation but on other principles. The commenter recommended an approach that ensures that weights depend on how much of the population is impacted by each measure and is based upon comparing hospital performance against best performers on an efficient frontier.
- One commenter stated data grouping should be consistent and not reported in homogenous grouping methods.
- One commenter added that grouping measures is a reasonable way to aggregate data. They added they are concerned with how measures have been assigned to measure groups and how the model has been validated for each measure group.
- One commenter noted insufficient information was released by CMS for stakeholders to evaluate the proposed changes to improve the model fit. The commenter noted that based on the limited information provided, the proposed changes are inadequate for improvement of the model.

Fifty-two commenters provided specific comments on the PSI-90 measure, including its inclusion within the Safety of Care measure group, and use of PSI components in place of the composite measure.

- Twenty-one commented that Safety of Care measure group scores are driven by the PSI-90 measure, leading to less evidence or review of performance on other meaningful measures in the measure group. They encouraged CMS to address the unequal loading of PSI-90 within the Safety of Care measure group. One commenter specifically mentions that the PSI-90 measure is easily gameable by hospitals.
- Nine commenters expressed concern that PSI-90 penalizes larger hospitals that perform more surgeries while being susceptible to surveillance bias (for example, some institutions may be penalized because they more diligently report safety events) and, as a claims-based measure, cannot capture the full scope of patient risk.
- One commenter stated, as a low-volume claims-based measure, it does not allow for patient-level risk factor adjustments.
- Two other commenters recommended removing all PSI measures from Star Ratings. Commenters expressed concern with the validity of the measures, noting they were developed to serve as flags of potential safety events, not as definitive performance measures.

- One commenter also stated that studies suggest the measures are insensitive and capture a small non-representative sample.
- One commented specifically how the PSI-90 measure in combination with the Hip/Knee Complication measure have historically dominated the Safety of Care measure group score. The commenter believed the HAI infection measures represent more important quality issues.
- One commenter stated the measure loadings within the Safety of Care measure group have not been steady over reporting periods and that PSI-90 has a disproportional impact on the measure group score.
- One commenter stated that the variance in surgical volume may disadvantage larger facilities, as larger institutions may have more complications in part to low volumes of cases at smaller institutions. The commenter added that there are also concerns about the reporting integrity at hospitals, and rewarding facilities for not accurately and transparently reporting events.
- One commenter supported PSI-90 being a focus point for safety measures. The commenter added that the PSI-90 measure includes more than just six metrics.
- One commenter supported the use of PSI-90 components.
 - The commenter suggested separating out the PSI components and testing each individual measure on its own merit.
- Ten commenters noted concerns with incorporating PSI-90 components as an alternative to the composite:
 - Two commenters expressed support for including measures of care that reflect and align with CMS's priorities, but added that using the PSI component measures instead of PSI-90 would be challenging. They added they agreed with the TEP to not support the regrouping options.
 - One commenter requested that CMS provide information about the statistical significance of the measured rates of individual PSI measures at the hospital level. The commenter added that denominators may vary drastically across hospitals and could exacerbate the biases seen in the PSI-90 composite. The commenter added that they do not support using the components in lieu of the PSI-90 measure and recommended the PSI-90 measure be removed from Star Ratings.
 - One commenter did not support the use of composite measures as the data is not conducive to quality improvement.
 - One commenter did not support use PSI-90 and requested further analyses on how measure grouping would account for low predictive values if PSI-90 component measures were used.

Ten commenters did not believe that regrouping would address underlying concerns with the existing methodology.

- One commenter said they did not feel the measure group changes will improve the sensitivity of Star Ratings and added that changes will not impact scoring unless the LVM methodology is addressed.
- Two commenters supported further analysis on how to improve measure groupings before implementing any changes.
- One commenter added that they would need stability, model fit, and confirmatory factor analyses in order to make a recommendation about Safety Of Care measure groupings.

- One commenter recommended CMS implement a simpler approach that focuses on consistent and balanced measure loadings.
- One commenter recommended keeping the current Safety of Care group with further exploration into the PSI-90 weighting, while retaining its composite.
- Three commenters suggested further refinement of the measure grouping and weights in a conceptually meaningful way that achieves measure loading balance.
 - One commenter felt the new approaches continue to show inconsistent imbalanced measure loading coefficients.

Two commenters noted the importance of balanced and consistent loadings (criterion 2) as it relates to regrouping options.

- One commenter added that such an approach would increase interpretability and add balance. The commenter urged CMS to consider simpler alternative approaches before implementing any regrouping.
- One commenter supported balanced and consistent loadings to evaluate measure groups but did not support any of the alternatives to grouping Safety of Care since they all resulted in unbalanced measure loadings, possibly due to the LVM.

Four commenters requested further information on measure grouping.

- Two commenters requested further input from stakeholders to conceptually and empirically evaluate the impact of regrouping and new group weights.
 - One commenter proposed only NQF-endorsed measures that are valid, reliable, and align with other measures be used in the methodology.
- Two commenters requested further exploration on the impact of measure regrouping and removal.

Response: Thank you for your comments. CMS will continue to reevaluate the manner in which measures are grouped within Star Ratings and the effect of adding and removing *Hospital Compare* measures.

The three-step approach of ensuring clinical groupings, conducting confirmatory factor analyses, and monitoring the balance and consistency of measure loadings has largely been accepted by stakeholders as an effective step in evaluating and maintaining measure groups while using LVM to calculate measure group scores. As noted by commenters above, alternative approaches to LVM may not require as rigorous criteria since policy-oriented approaches to aggregating data that do not utilize statistical models may not require the same technical assumptions of the LVM.

We agree that the alternative approaches to grouping Safety of Care, including separating into surgical vs medical and using PSI components in place of PSI-90 composite, did not result in favorable confirmatory factor analyses or balanced and consistent measure loadings. However, we wanted to be transparent about CMS efforts to address stakeholder concerns regarding measure loadings, particularly PSI-90, within Safety of Care.

As part of the meaningful measure initiative, CMS had previously announced intent to and recently began to retire measures from *Hospital Compare* to streamline measurement and reduce hospital burden. CMS will continue to monitor the available measures.

Incorporating Measure Precision

Twenty-seven commenters agreed with incorporating measure precision within Star Ratings.

- Six commenters supported seeking a general alternative method to incorporating measure precision.
 - Four commenters agreed that measure precision where possible should be considered but noted that no strategy has a definite advantage over another. They noted that regardless of the method adopted, transparency would assist stakeholders with identifying accurate calculations.
 - One commenter advocated for increased precision and accuracy over the existing model.
 - One commenter said that they generally support any statistical improvements that increase precision and decrease bias or shifts in Star Ratings.
- Five commenters supported the current denominator weighting approach to incorporating measure precision, weighting measures based on hospital volume.
 - One commenter noted denominator weighting is more transparent and can be applied more consistently across measures.
 - One commenter encouraged CMS to consider inclusion of confidence intervals and measure weighting as part of the strategy.
 - One commenter noted their support on for denominator weighting, particularly for HAI measures.
- Ten commenters supported confidence interval weighting as an alternative to the current approach of denominator weighting:
 - One commenter recommended this approach be applied to measure groups with imbalanced measure loadings.
 - One commenter noted confidence interval weighting results in the most balanced measure loadings of all the options.
 - One commenter supported confidence interval weighting and limiting the impact of volume adjustment.
 - One commenter supported the use of either log transformation and/or confidence interval weighting.
 - Three commenters suggested that the shift to confidence interval weighting would help ensure more balanced measure loadings within the Safety of Care measure group.
- Four commenters supported log transformation of the denominator.
 - Two commenters recommended a natural-log based scaling approach and explained that this reduces the complexity. One encouraged CMS to consider normalizing the log-transformed denominators with min/max scaling.
 - One commenter explained that log transformation helps with normalization of data when outliers are present.
 - One commenter suggested log transformations in the denominator for more equitable distribution of loadings.

- One commenter supports log transformation and notes the benefits of improved consistency and distribution of measure loadings. However, they advocated for equal measure weights as part a long-term update since it would be the most intuitive approach to weighting measures.
- Eight commenters supported implementing a combined approach to incorporating measure precision.
 - Another commenter supported combination methodology and advised to use weighting method that best suits the data in a given measure group.
 - One commenter suggests applying a uniform weighting method across all measures. They also suggest a change in weighting method be applied only to measure groups that meet specific criteria.
 - One commenter supported confidence interval weighting and log transformation as a combined approach to incorporating measure precision.
 - One commenter noted that using a combination of denominator weighting and log (denominator) weighting would be the ideal balance of measure loadings within measure groups.
 - One commenter noted that using a combination of denominator weighting and log (denominator) weighting would be an improvement, but a better improvement would be using the explicit approach based on relative clinical significance. The commenter believes that weighting measures by precision creates importance for events that may not be clinically important, such as accidental lacerations that patients may not even notice.
 - One suggested greater uniformity of measure loadings and supported all options set forth, either no weighting, denominator weighting, log weighting, or confidence interval weighting.
 - One commenter noted they have no specific concerns regarding a combination approach of denominator and log denominator weighting.
 - One commenter noted that log weighting is complex and supported the applying different measure precision approaches across all measure groups, which may be necessary with different types of measures.
- One commenter discourages a mixed weighting methodology and requests CMS to be transparent with stakeholders regarding any changes in future methodology.
- One commenter noted that measure precision is critical to the Star Ratings methodology. However, they expressed concern that CMS and stakeholders lack of holistic understanding of the different measure precision options and recommends further analyses before any methodology updates are implemented.

Response: Thank you for your comments and support. We agree that accounting for precision is important. Some, but not all, of the measures included within Star Ratings already account for measure precision within the individual measure methodologies.

The measure loadings produced by the LVM consider both hospital volume and the correlation between the measures in a measure group with the underlying trait of quality (that is, the measure group score). For this reason, measures that are both highly correlated with the measure group score and have large

denominators, such as PSI-90 within Safety of Care and Hospital-Wide Readmission within Readmission, have high measure loadings when using the denominator-based weighting approach (the current methodology). Other weighting approaches (such as no weight) result in more balanced and consistent loadings, but do not account for measure precision to the same degree or across measure groups.

We appreciate comments generally supporting the principle of acknowledging differences in measure score precision between hospitals as well as acknowledging the technical challenges of each proposed approach given limitations in measure precision data availability. For example, hospital-specific standard errors are not available for all measures, therefore we are exploring several alternatives based on available data.

Eight commenters specifically did not support the current approach of denominator weighting.

- One commenter noted that the assumption of large volume being equivalent to increased measure precision is a logic error. If there is not enough volume to make the data for the measure statistically valid, then the measure should not be included or reported. They noted that artificially determining measure precision based on volume is not a rational method.
- One commenter stated that incorporating measure precision at the individual measure level creates unintended consequences of unevenly distributing star ratings based on hospital size. The commenter stated that the volume adjustment in LVM causes small hospitals to receive middle star ratings and large hospitals to receive high or low star ratings.
- One commenter encourages the change of the current use of denominator weighting and expresses concern about the accuracy of this method. They note that some clinical conditions have higher volumes than others but should not be weighed more heavily in an overall score.
- One commenter notes the ongoing concern for accuracy in capturing outcomes when utilizing denominator weighting.
- One commenter stated that the methodology should account for the reliability of each measure but noted that it was unclear if measure loadings reflect quality or hospital volume.
- Two commenters do not support noted that the current approach of denominator weighting and noted denominator weighting contributed to the poor balance of measure loadings in the Safety of Care measure group.
- One commenter specifically did not support using confidence intervals as an approach to incorporating measure precision since accounting for varying populations across measures would be difficult with varying confidence intervals.

Twelve commenters expressed concerns about LVM as it relates to incorporating measure precision

- One commenter expressed concern about the latent variable approach and explains that the incorporation of measure precision makes the ratings less reliable for all hospitals. They suggest increasing the number of measures for more balance and stability.
- One commenter noted that LVM is not appropriate and mentioned that recent changes to use alternative HAI measure denominators for measure precision has had no significant impact on ensuring a consistent model.
- Another commented that the choice of confidence intervals, as well as log transformations, where both produce radically different loadings from one another

further enable an opportunity to question the LVM approach. They encourage CMS to consider how the methodology will impact the reality of healthcare.

- Eight commenters attributed imbalanced measure loadings to the incorporation of measure precision.
 - One commenter noted that higher measure loadings have a sensitive impact on overall scores and do not provide reliable/accurate information.
 - Four commenters expressed concern that incorporating precision would further imbalance measure loadings within the Safety of Care measure group.
 - Three commenters noted incorporating measure precisions creates an imbalance of measure loadings for PSI-90 and other underlying measures. One commenter encouraged CMS to address the imbalanced measure loading of PSI-90 within the Safety of Care measure group.

Response: Thank you for your comments and feedback. CMS will consider your comments. While the current use of measure denominators to account for measure precision is partially responsible for imbalanced measure loadings, measure precision is important in ensuring that measures with more patients and information contribute more to a hospital's measure group score. We agree with comments from statistical experts that we can be more certain about results from measures with more information. While using hospital-specific measure score standard errors would be ideal for incorporating measure precision information into the LVM, these data are not currently available for all measures and have necessitated the use of denominator weighting in the current methodology. As CMS considers updates to the methodology, data availability, face validity, and technical considerations will all be considered to generate solutions that are both feasible and meaningful.

Period-to-Period Shifts

Overall Period-to-Period Shifts

Thirteen commenters emphasized the importance of addressing period-to-period shifts in the Overall Star Rating methodology.

- Six commenters noted stability in star ratings is important because providers use Star Ratings to drive quality improvement.
 - Three of the commenters stated significant shifts in ratings between measurement periods may cause the ratings to appear random, and deter hospitals from quality improvement activities.
 - One commenter asked CMS to actively follow measure groupings for consistency in how each measure influences the group score over time.
- Four commenters noted that stability in Star Ratings is important because consumers use star ratings to make healthcare decisions.
- One commenter stated that significant period-to-period shifts can be confusing for the general public, who may not understand the Overall Star Rating methodology and, thus, the reasons for shifts.
- One commenter noted the period-to-period shifts question the validity of the methodology.
- Two commenters stated it can be hard for hospitals to explain large shifts in their star ratings despite relatively modest changes in performance on individual measures.

- One commenter noted changes in star ratings across reporting periods should reflect significant changes in measure performance, rather than instability in the Star Rating methodology. Since Star Ratings are publicly reported, hospitals should be able to clearly see how changes in their underlying measure performance affects their star rating.

Twelve commenters urged CMS to review and update the current methodology to improve the predictability and consistency of the Overall Star Rating across periods.

- One commenter noted low patient volume also decreases predictability of the Star Ratings, which negatively affects CAHs. This commenter urged CMS to resolve these issues for smaller hospitals and improve predictability.
- One commenter urged CMS to review causes of significant shifts that have occurred in recent updates to determine whether changes to the methodology are needed.
- One commenter suggested providing a transition period for any significant changes to the Overall Star Rating methodology to give hospitals time to make adjustments if needed.
- One commenter attributed period-to-period shifts in star ratings to changes in measure-level performance and recommended any measure with volatility from year to year be removed from the Star Rating due to lack of reliability.
- One commenter suggested that peer grouping may minimize period-to-period shifts.
- One commenter recommended achieving consistency by using the weighted average over the same period for measures with no overlapping data as a way to mitigate any adverse impact from measures using data from different time periods.
- One commenter suggested that CMS move to a 1-3-5 rating system to simplify the ratings and reflect the worse than, same as, and better than categories already displayed on *Hospital Compare*. They believe this update combined with an explicit approach would increase face validity and reduce shifts in star rating categories for hospitals on the edge of star clusters.

Five commenters provided additional thoughts related to period-to-period shifts.

- Two commenters acknowledged that shifts in star ratings are expected, except for shifts in two or more star categories that is driven by random fluctuation rather than hospital performance.
- One commenter noted with the current methodology it is unclear how hospital performance on measures and improvement efforts impact their star rating.
- One commenter stated hospitals generally cannot predict how their performance on the underlying measures will translate into a star rating, which could disincentive any quality improvement efforts.
- One commenter expressed concerns about stability of reporting the Star Ratings. This commenter noted the Overall Star Rating was not updated for nearly 1.5 years due to concerns with the scoring methodology, resulting in consumers receiving inaccurate information and hospitals experiencing changes in their scores due to measure retirement and additions during this time.

Response: Thank you for your comments. Previous analyses have revealed that the current methodology is sensitive to changes in the underlying measures. While some shifts in star ratings are expected as hospital performance changes from period to period and a specific hospital's performance changes in

relation to all other hospitals in the nation, we agree that extreme shifts in star ratings can be confusing to hospitals and consumers. Several of the methodology updates outlined in these materials are aimed at reducing extreme shifts in Star Ratings over time.

Annual Reporting

Forty-two commenters supported an annual refresh of the Overall Star Ratings. Of the commenters who expressed support, some noted additional thoughts.

- Fourteen commenters supported an annual refresh with the rationale that some measures are only updated annually (such as the readmission and mortality measures), while other measures are updated more frequently on *Hospital Compare*.
- Two commenters stated the refresh should occur in July of each year.
- Two commenters supported an annual refresh but also suggested providing confidential performance reports every four or six months to help providers refine their improvement initiatives.
- One commenter supported annual reporting alignment with other quality and incentive program reporting as well as other CMS quality rating systems, such as Medicare Advantage and Part D Star Ratings.
- Two commenters stated an annual refresh could demonstrate consistency and eliminate any consumer confusion.
 - One commenter noted that a consumer could be confused if they had chosen a facility based on a rating and that rating dramatically shifted within a couple of months.
 - One other commenter pointed out that most healthcare organizations are already tracking measure performance on a monthly basis for improvement.
- One commenter noted an annual refresh would allow changes in a hospital's star rating to be more clearly attributed to observed changes in the hospital's performance for individual measures.
- One commenter stated an annual refresh may better allow for measure-level changes and mitigate the impact of measure periods with incomplete data on the Overall Star Ratings.
- One commenter noted small rural hospitals have smaller populations, and may benefit from an annual refresh that will help minimize larger than expected shifts in star ratings.
- One commenter stated an annual refresh would allow sufficient opportunities for stakeholder and public feedback between cycles.
- One commenter noted an annual refresh would allow a longer lead time to address issues identified by hospitals and more easily implement recommendations to improve the accuracy and reliability of the Overall Star Ratings.
- Two commenters stated their support of an annual refresh is contingent upon methodology updates.
 - One commenter specifically requested revisions to the LVM approach.
 - One commenter specifically requested inclusion of meaningful measures, defined measure weights, and the application of appropriate peer grouping.
- Two commenters supported an annual refresh of Star Ratings as a short-term solution.
 - One commenter added if the annually reported measures were updated more frequently, they would support a biannual update of the Overall Star Ratings.

- One commenter added once the model is shown to be reliable and reproducible, refreshing star ratings more often could provide value to patients and providers.
- One commenter noted a drawback to annual refresh is it limits visibility of improving and worsening of scores for measures that are updated multiple times a year. Nevertheless, this lag would be no longer than nine months in the worst case.
- One commenter suggested that CMS host a confidential preview of potential methodology updates concurrent with the public comment period.
- One commenter stated it is important that CMS and hospitals have adequate time to understand the proposed methodology updates and ensure that it does not disadvantage different hospital types before the next release of star ratings.

Thirteen commenters supported continuing to refresh the Overall Star Ratings biannually and expressed concerns about transitioning to an annual refresh.

- Six commenters stated patients and consumers should be given the most relevant and timely information available to make decisions about their healthcare.
- Two commenters noted a biannual refresh gives hospitals the opportunity to improve their score more than once a year.
- One commenter noted a biannual refresh ensures alignment with more frequent updates in HAI (within Safety of Care measure group) and Hospital Consumer Assessment of Health Providers and Systems (HCAHPS) (within Patient Experience measure group) data. This commenter added an annual update could mask incremental improvements made by hospitals on these measures.
- One commenter added that refreshing the Overall Star Ratings less frequently penalizes hospitals working to improve their quality.
- One commenter stated an annual refresh will decrease face validity for hospitals and consumers, and the star ratings will not coincide exactly with the data and outcomes reported for measures on *Hospital Compare*.
- One commenter noted CMS should not take actions to limit the shifts, as shifts in the Overall Star Ratings may reflect other factors that need to be evaluated and addressed in the methodology.
- One commenter noted other Star Ratings systems, such as Nursing Home Compare, refresh ratings even more frequently (on a quarterly basis).
- One commenter noted annual release of Star Ratings would not address the model's current flaws.

Three commenters provided other thoughts on refreshing the Overall Star Ratings.

- One commenter stated that star ratings should not be refreshed more than twice each year.
- One commenter suggested refreshing star ratings quarterly, since facilitates submit data quarterly.
- One commenter suggested the Star Ratings should be released on a regular and consistent timeframe, and highlight the measure performance time period for consumers. This commenter added CMS should help patients and consumers understand that there is a lag in the performance period for the Overall Star Ratings, so shifts between six-month periods do not reflect current improvements or worsening in quality.

Response: Thank you for your comments. Previous analyses have revealed that there are a number of hospitals with summary scores at the borders of k-means clusters, used to assign star ratings, that may be more prone to shifting between star rating categories each reporting period. Based on historical data, we anticipate that an annual refresh of the star ratings will minimize these quarterly or semi-annual shifts in star ratings, and in turn improve the face validity of the ratings. We have also heard that hospitals shifting star rating categories may encounter difficulties explaining the shift in star rating to hospital leadership and consumers on a quarterly or biannual basis. We hope an annual refresh would reduce provider burden and make each release of star ratings more manageable and easy to interpret. An annual refresh also carries the additional benefit of affording CMS sufficient time to validate individual measure and star rating results prior to confidential reporting for hospitals and public reporting.

Weighted Average Summary Scores

Four commenters supported modifying the Overall Star Rating methodology to incorporate data from previous periods through a weighted or time averaged approach.

- One commenter suggested using a weighting of 25% old and 75% new data. This commenter noted CMS should carefully consider how much historical data to incorporate into the score, since it could anchor the overall score depending on the weighting.
- One commenter alternatively suggested using a two-year lookback period for all measures to provide hospitals and consumers with more current data and demonstrate a hospital's improvement sooner.

Twenty-four commenters did not support a weighted or time averaged approach.

- Six commenters noted any quality ratings system should reflect the most current quality information to the greatest extent possible. Incorporating previous performance would not be an accurate reflection of hospitals' current performance.
 - One commenter pointed out that the TEP and work groups shared the same opinion.
 - Another commenter stated that measure performance can change and those changes should be reflected in the star ratings. While consistency and transparency are important, these goals shouldn't be achieved through use of past performance data.
- Thirteen commenters expressed concern that the incorporation of historical data within the Star Ratings would obscure or hide improvements in measure performance.
 - One commenter specifically mentioned that the historical data may disincentive hospital improvement efforts.
 - While the commenter did not support a weighted average at this time, one commenter noted that as calculation and reporting methods allow inclusion of more timely data in the future, a 75% new and 25% historical data method would smooth period-to-period shifts.
- Five commenters expressed concerns about incorporating historical data into the Star Rating because it may be misleading to consumers who would be making decisions based on outdated information.
- Four commenters noted some heavily weighted measures, such as the readmission and mortality measures, are already based on multiple years of data. Incorporation of previous

performance periods in addition to multi-year measure data would further limit the usability of the Star Ratings.

- One commenter noted incorporating historical data within the Star Ratings seems cumbersome and error prone and could lead to unintended consequences.
- One commenter suggested publishing both prior and current Star Ratings.
- One commenter suggested exploring other changes in the methodology instead, such as using confidence intervals to account for measure precision or removal of the PSI-90 composite measure from the Safety of Care measure group.

Five commenters suggested exploring the use of partial or half star ratings to address period-to-period shifts.

- One commenter noted this approach would reduce the “cliffs” between hospital categories and provide greater clarity to patients and consumers on hospital performance.
- One commenter noted the Home Health Compare Star Ratings and the Medicare Advantage Plan Quality Star Ratings programs already use half stars.
- Two commenters supported using half stars to reduce period-to-period shifts.
 - One commenter noted half stars could help distinguish between higher and lower performers within a given star rating category.
 - One commenter noted, if considered, empiric evaluation and consumer testing of half star ratings would be needed. This commenter also supported using three star categories, which would provide patients with useful information on outliers and improve period-to-period consistency.

Fourteen commenters proposed using more recent data to derive the Overall Star Ratings, attributing the shifts in Star Ratings to refreshing measures that use multiple years of data.

- One commenter stated that heavily weighted outcome groups use outdated data, which does not reflect current performance, is misleading to the public, and limits the usefulness of Star Ratings.
- Three commenters suggested using more recent data for outcome measures with three-year performance periods to be more reflective of a hospital’s current performance.
 - One of the commenters stated this will ensure consistency with the measure periods in other measure groups, such as process measure groups, and better reflect current hospital quality.
- An additional commenter noted that measures that use older data make it difficult for improvements to be reflected in scores.
- Another commenter noted the current Star Rating methodology reflects hospital performance data no more recent than one year and as far back as three years, which is not as valuable to consumers. In addition, since it takes two or three years for past performance to not be included in the current star rating, a hospital could have a very high or very low star rating that has no correlation to their current performance.
- One commenter explained that the use of older, poorer performance data results in a negative hospital image by the public, patients, competitors, insurance companies, and the media and forces hospitals to provide additional proof and data that reflect more current,

improved metrics. The commenter suggested including a disclaimer on *Hospital Compare* about the data collection period.

- Two commenters noted the performance periods of individual measures vary, and suggested CMS align measure timeframes to promote transparency and reduce provider burden. One commenter noted multiple measure timeframes is suboptimal from the hospital and consumer perspectives.
- Three commenters recommended using data no older than six months for star ratings.
 - One commenter added since the data collection period is so long, it can take 1-2 years for star ratings to reflect hospital improvement efforts.
 - One commenter suggested a rolling quarter update, which would help hospitals make improvements in a timelier manner.
- Another commenter suggested using real-time data.
- One commenter noted significant shifts in star rating suggest outdated data is being used to determine the ratings.

Response: Thank you for your thoughtful comments and suggestions on addressing period-to-period shifts. We agree that some changes in star ratings are expected and should reflect changes in hospital performance and quality improvement efforts. We also agree that shifts by two or more star ratings categories should be rare and only occur in response to drastic changes in quality measure availability and performance. The TEP, Provider Leadership Work Group, and Patient & Patient Advocate Work Group convened by CORE also agreed that it was more important to use current data rather than include older data to introduce more consistency in scores. We will explore other methods, including some of the potential updates outlined within these materials, to minimize extreme shifts in star ratings in conjunction with other methodology changes.

Peer Grouping

Eighty-five commenters supported peer grouping.

- Eighteen commenters supported peer grouping based on differences in hospital case mix or service offerings.
 - One commenter recommended exploring peer grouping options that allow hospitals with similar characteristics and risk profiles to be compared to one another. The commenter recommended using vast resources and data to determine the most appropriate peer groups, and sharing the methodology with hospitals in advance of implementation.
 - Two commenters supported peer grouping in the short term stating that teaching hospitals perform a wide array of complicated and common procedures, new treatments, and care for broader socio-demographic patient populations yet are compared to hospitals with limited patient populations and procedures performed, which disadvantages large teaching hospitals. The commenter recommended that CMS explore measure performance within specific hospital peer cohorts so that hospitals with similar characteristics and risk profiles are compared to each other. The commenter suggested looking at other star ratings that account for differences in hospitals (such as Nursing Home Compare Star Ratings) to inform the methods. The

commenter noted one suggested variable could be hospital size/full service status, to ensure patients are able to compare hospitals that are able to fully meet their care needs.

- One commenter noted that hospital size, location, and specialties all influence the types of patients for whom a hospital provides care. They added volume adjustments in the underlying measures cause biases in star rating distributions, and grouping by hospital size can help reduce these biases.
- One commenter stated that hospitals do not all provide the same complexity or breadth of care, so Star Ratings make more sense when assigned based on the type of care and services the hospital provide. The commenter added that patients are seeking care for a specific condition, and not all hospitals care for that condition.
- One commenter noted that facilities with small population size and offer limited services receive higher ratings but, in reality, those hospitals are transferring complex patients to other hospitals that do offer broader services. The commenter recommended that the methodology account for complexity of care or services provided.
- One commenter stated that hospitals vary in services and acuity of patients, so differences in hospitals and factors outside of hospitals' control should be accounted for. The commenter recommended peer grouping by type of hospital (specialty, teaching) as an interim step to true risk adjustment. The commenter added that CMS should examine this approach, with input from stakeholders, to identify both the variables by which to peer group and the usefulness to the patient in having this information.
- One commenter stated that currently hospitals with more complex care and more vulnerable patients are unfairly penalized. A Quality Accountability framework could be utilized to cohort hospitals based on volume thresholds (solid organ transplants, acute transfers-in, trauma cases, cardiothoracic and neurosurgery volumes) that differentiate patient complexity. This would create meaningful and actionable benchmarks and comparisons.
 - The commenter added that of the proposed peer grouping variables (dual eligibility, teaching status, number of beds), no one single factor or characteristic provides sufficient separations or adjustments for hospital differences. Number of measures reported is a step in the right direction towards volume of patients seen.
- One commenter stated that differences in hospital characteristics, such as teaching or safety-net status, or range of services provided are not accounted for. Academic teaching hospitals perform a wide array of procedures and care for broader socio-demographic patient populations compared to other hospital types, leading to a disadvantage in the Star Ratings methodology for large teaching hospitals. The commenter added that a recent study showed ratings do not fully represent the risk of undergoing procedures at low volume hospitals.
- One commenter recommended selecting a peer grouping approach that allowed for like to like comparisons. The commenter requested considering past experience with the Emergency Department Throughput Measures. The commenter noted if that was not

possible, CMS should institute a severity adjusted, hospital type, patient type and volume like to like comparison.

- The commenter added that many measures included in Star Ratings unfairly impact teaching hospitals that treat low socio-economic status patients, and that CMS has recognized the need for socio-economic status adjustment in many other programs.
- One commenter recommended rating hospitals in comparison to like facilities, similar to categories used in a public hospital report card, Watson Health, developed by a private company. The same commenter recommended accounting for social determinants as they relate to the challenges faced by Safety Net hospitals.
- One commenter recommended stratification similar to Hospital Readmission Reduction Program (HRRP) quintiles, but not by the proportion of dual-eligible patients. The same commenter noted disparity between major teaching hospitals that treat complex conditions for patients with socio-economic risk factors, and specialty hospitals that work with insured patients on elective procedures. They further stated that 52.5% of the 40 four- and five-star major teaching hospitals performed below average on readmissions, pointing to inequality in patient frailty.
- One commenter noted 61% of specialty hospitals received a 5-star compared to 9% of major teaching hospitals. Major teaching hospitals provide a vast array of services to a diverse payor mix compared to small specialty hospitals.
- One commenter stated that star rating performance across hospitals demonstrates bias towards lower ratings for teaching hospitals and safety net hospitals, and higher ratings for specialty hospitals and those reporting fewer measures, which should be accounted for in peer grouping.
- One commenter further noted that small community hospitals with minimal services shouldn't have a higher rating than academic medical centers and safety net hospitals and that the Star Ratings do not help consumers make decisions about life-threatening conditions.
- One commenter stated that the Star Ratings do not serve the intended goals for small, rural, and CAHs, and hurt the providers who rely on relationships with their communities. The commenter added that Star Ratings negatively impact consumer ability to make educated care decisions.
- One commenter stated that small hospitals often do not have enough measures to qualify for some measure groups, consequently putting higher weight on other measure groups.
- One commenter also suggested investigating additional strategies to make it easier for health systems to understand recent performance of peer institutions on clinical outcomes contained in the Star Ratings.
- Three commenters proposed adjustment for socioeconomic status, as they affect health outcomes. The commenter noted not adjusting for social determinants of health will put certain hospitals at a disadvantage.
 - One commenter noted they support peer grouping as a means to address potential bias.

- Two commenters felt there was a risk for large hospitals, teaching hospitals, and hospitals serving low-income patients to receive a lower star rating despite providing quality care.
 - One of the commenters stated sound data to reduce disparities in care is important.

Sixty-one commenters proposed a variety of peer grouping suggestions without specific variables.

- Twenty-five commenters recommended peer grouping by teaching status, CAHs, and proportion dual-eligible patients given the strong impact of socio-demographic factors on patient outcomes.
- One commenter supported peer grouping of hospitals by complexity of patients seen and services provided.
- One commenter recommended displaying peer groups for academic/teaching and non-academic/non-teaching, rural, community for-profit, community non-profit, critical access designation, and bed size categories of <50, 51-99, 100-299, 300-499, and >500.
- One commenter recommended peer grouping by teaching status, facility, size, or socioeconomic status.
- One commenter suggested peer grouping by size, geography and/or teaching status.
- One commenter suggested peer grouping by safety net hospitals, percent Medicaid, type of community, and region.
- One commenter recommended peer grouping including adjustments for size, volume, socioeconomic status impact, services provided and not provided, and measure submissions/exclusion of measure.
- One commenter proposed adopting the risk-stratification to group hospitals approach used in HRRP for peer grouping.
- One commenter recommended stratification by sociodemographic factors (dual eligible, area income statistics), teaching status, and provider status (CAH vs. Prospective Payment System) to help account for limitations in specific domains due to reporting restrictions.
- One commenter recommended peer grouping by specialty. The commenter noted that bed size may also serve as a proxy for types of services provided and similar facilities. The commenter added that although teaching status has been suggested as a proxy for social disparities, many non-teaching hospitals also care for this patient population.
- One commenter recommended peer grouping by dual eligible status, type of hospitals, number of measures reported, and CAHs to address hospital characteristics, as hospitals who report fewer measures receive higher star ratings.
- One commenter recommended grouping large teaching hospitals, community hospitals, by bed size, and by number of measures reported. The commenter added that it is unfair to compare hospitals that report many measures with hospitals that report few measures.
- One commenter recommended peer grouping by teaching status, distinguishing major and minor teaching hospitals, and grouping CAHs together. The commenter also recommended not including single specialty hospitals in Star Ratings.
 - The commenter did not support peer grouping by number of measures reported, as the underlying assumption of Star Ratings that as many hospitals as possible should be included is flawed. The commenter added that grouping hospitals by number of

measures reported is not a meaningful peer group because the hospitals could be reporting different measures.

- One commenter recommended including teaching status, measures, bed count, and critical access status.
- Since perfect risk adjustment is not possible, the commenter recommended peer grouping by teaching status, size, or socioeconomic status.
- One commenter recommended peer grouping by teaching hospitals, CAHs, rural hospitals, and community hospitals to account for the differences in services provided among hospitals.
- One commenter suggested peer grouping by CAHs/hospitals with less than 100 beds, specialty hospitals, large academic/teaching hospitals, and other medical surgical hospitals.
- Another commenter recommended peer grouping by hospital bed size, Prospective Payment System Hospitals verse CAHs, patient complexity, number of measures reported per measure group, for profit verse not-for-profit status, and community verse private owned to account for differences in hospital types.
- One commenter recommended three groupings: comprehensive academic medical center, community hospitals, and complex care medical center group. The commenter also suggested a separate framework for CAHs and specialty hospitals.
- One commenter recommended accounting for proportion of uncompensated care or uninsured patients, as safety net hospitals provide essential access to health care to this population. The commenter added that safety net hospitals typically serve patients of higher acuity, and recommended grouping hospitals by acuity based on case-mix index.
- One commenter recommended stratification by teaching status, CAHs, urban/rural status, disproportionate share hospitals, case-mix index, patient volume variables, and proportion of dual-eligible patients.
- One commenter recommended peer grouping by case mix index, trauma service, acute transfers in, socioeconomic status, and total discharges in order to compare like to like hospitals.
- One commenter suggested using case mix index and number of beds to achieve grouping like to like hospitals in a simple way that patients could understand.
- One commenter noted small hospitals more often receive 5-stars. The commenter recommended peer grouping by hospital size, location, specialty status and teaching designation, which all influence types of patients a hospital cares for. This would provide a more meaningful framework for consumers and be more fair to the hospital providers.
- One commenter stated large medical facilities should not be grouped with small critical access facilities.
- One commenter noted the differences in complexity of care provided at academic medical centers and rural community hospitals. The commenter recommended peer grouping by the Vizient Quality and Accountability Study cohorts: comprehensive academic medical centers, complex teaching hospitals, and community hospitals, as well as a separate cohort for safety-net hospitals. The commenter also recommended grouping based on payer mix.
- One commenter noted peer grouping by 1) proportion of dual-eligible patients would be plausible and align with HRRP; 2) teaching status would also be reasonable since other rating systems, such as US News, do separate hospitals by teaching status but pointed out the categories would be arbitrary and CMS would need rationale for using teaching status instead

of Safety-Net designation; 3) bed size would not be useful since there is not evident link between bed size and star rating performance; 4) number of measures would be the best variable considering the relationship between number of measures reported and star ratings and the association to hospital type, however cut-points would need to be determined; 5) rural vs urban hospitals could be intuitive but would not facilitate consumer use since consumers in rural areas would not have many other hospitals for decision-making; 6) specialty designation would be conceptually useful.

- One commenter recommended focusing on the volume of services and patient conditions cared for.
- One noted eight characteristics which are significant in identifying distinct hospital cohorts: total outpatient visits, acute transfer volume, case mix index, inpatient surgical cases (as percent of all admissions), outpatient surgical cases (as percent of total surgical cases), trauma services, bone marrow transplant services, solid organ transplant services.
- One commenter recommended grouping hospitals, including removing critical access and specialty hospitals, and found that academic medical centers were more evenly distributed among 4 and 5 stars and complex and community teaching hospitals did not change star ratings significantly. The commenter stated this approach is more practical and provides a fair assessment of hospital performance.
- One commenter noted that if peer grouping is not feasible, risk adjustment for facility size or socioeconomic status should be considered.
- Seven commenters suggested peer grouping by proportion of dual-eligible patients, in alignment with the methods used for the HRRP.
 - One commenter's analyses showed a statistically significant correlation between Star Rating summary scores and proportion of dual-eligible patients.
 - One commenter stated this approach would more fairly compare performance among hospitals and lead to more accurate star ratings.
 - Another commenter noted that numerous studies show that social determinants of health contribute to mortality and readmission, so should be accounted with in Star Ratings.
 - One commenter noted that dual-eligible beneficiaries represent specific needs within a community and result in a great amount of care provided by hospitals.
 - One commenter noted the current discordance between programs is confusing for hospitals.
 - Another commenter stated that socio-economic status is adjusted for in HRRP but not in the Star Ratings readmission domain, creating inconsistency, and therefore should be adjusted for in Star Ratings.
- Two commenters recommended peer grouping by teaching status.
 - One commenter stated comparing similar hospital types will better inform consumers about what to expect from types of hospitals and will be less confusing, and recommended presenting all academic medical center Star Ratings together on *Hospital Compare*.
 - One commenter noted the methodology does not account for the complexity of care provided by academic medical centers.
- Two commenters recommended peer grouping by bed size.

- One of the commenters noted that hospital size can bias results and CMS risk-adjustment methods favors smaller hospitals.
- One commenter stated that bed size would help account for differences in services provided, size, patient population, and reported number of measures. The commenter added that a simplistic approach should be developed as star ratings are intended to be understood by consumers.
- One commenter noted bed size impacts resources available.
- One commenter recommended peer grouping by measure information, as small hospitals with limited number of measures should not be compared to large hospitals who report on all of the measures.
- Five commenters recommended peer grouping by rural or critical access designation.
 - One commenter noted the low volume metrics and participation in Medicare Beneficiary Quality Improvement Project (MBQIP) metrics but not Value Based Purchasing metrics.
 - One commenter noted many CAHs do not meet the minimum dataset requirements. The commenter added many CAH measures are outpatient measures, and patient compliance is difficult in these.
 - One commenter suggested grouping CAHs or bed size so that small hospitals were compared to similar hospitals.
 - One commenter suggested peer grouping by CAH designation. The same commenter noted that there should be an overall star rating and then star ratings based on the most common or costly procedures.
 - One commenter added that CAHs typically perform average in Safety of Care because they are excluded from the heavily weighted PSI-90 measure.
 - One commenter was neutral on peer grouping. The commenter noted it is possible to be a 5-star CAH, and that patients in CAHs should have the same quality of care that an urban hospital provides.
 - One commenter stated that the Star Rating methodology does not suit CAHs, which are not required to report the measures included in Star Ratings. Instead, CAHs report through MBQIP which has little overlap with measures included in Star Ratings. The commenter added that Star Ratings were therefore designed for large acute care hospitals.
 - The commenter added that the reporting threshold and methodology impacts smaller hospitals as group/measure weights are reassigned and greater emphasis is placed on measures that disproportionate impact hospitals with smaller patient volumes. Additionally, CAHs typically fall at the border line of having enough measures to be included in Star Ratings, and falling short of the reporting threshold. Often, this forces consumers to choose between facilities with a low star rating or hospitals who do not have enough information to receive a star rating, which is challenging.
 - The commenter recommended addressing small, rural, and CAH challenges by removing CAHs from star ratings or allowing CAHs to opt out of star ratings. The commenter stated this will help mitigate harm suffered by small hospitals,

provide accurate comparisons for patients, and mitigate fluctuations in star ratings for hospitals reporting few measures.

- Two commenters recommended peer grouping by specialty hospital designation.
 - One commenter provided rationale that the level of services, complexity and acuity, and types of patients seen are very different between general community hospitals and specialty hospitals.
 - One commenter suggested using the same comparative groups as Leapfrog.
 - One commenter suggested assigning hospitals to peer groupings based on size or status, noting it would be beneficial to both hospitals and patients. The commenter also noted that patients are generally not familiar with hospital designations.
 - Two commenters suggested general hospitals only be compared to general hospitals, but that specialty hospitals continue to be compared to all eligible hospitals.
 - Three commenters who recommended peer grouping for the short-term suggested variables that are generally outside of the control of hospitals, including number of reported measures and the proportion of dual-eligible patients. The commenters added that there are inherent shortcomings to the dual eligibility methodology used in HRRP.
 - Another commenter also supported a peer grouping methodology as an interim step towards risk adjustment.
 - Another commenter stated that hospitals differ by case mix and complexity and when hospitals are not appropriately stratified, it leads to unfair and potentially misrepresented performance comparisons. They suggested creating a peer group methodology and sharing it with hospitals for input.
 - One commenter suggested a risk-adjusted approach to normalize comparative data. For example, case mix index could be used as an indicator of case complexity. The commenter noted that trauma, cancer, and burn centers may have more readmissions, infections, and mortality events, for which the current methodology does not account.
 - One commenter also recommended not including single specialty hospitals in Star Ratings. The commenter did not support peer grouping by number of measures reported, as the underlying assumption of Star Ratings that as many hospitals as possible should be included is flawed. The commenter added that grouping hospitals by number of measures reported is not a meaningful peer group because the hospitals could be reporting different measures.

Fifteen commenters supported peer grouping as a short-term strategy but requested direct risk adjustment in the overall methodology for the future.

- Two commenters noted rating should account for differences in clinical and social risk factors, and hospitals that serve sicker and poorer patients should be on a level playing field with other hospitals. These hospitals tend to perform worse on Star Ratings, specifically teaching hospitals, hospitals that report more measures, and hospitals receiving the highest disproportionate share hospital (DSH) payments.
 - One commenter stated that peer grouping can be a short-term strategy to address biases, but direct risk adjustment approaches should be explored for the future as it would help improve the performance comparisons.

- One commenter added that more Massachusetts hospitals are slated to receive DSH payments in 2019 than the national average, showing that these hospitals care for a significant proportion of disadvantaged patients.
- One commenter stated that in the short term, a two-factor approach in LVM that includes one factor that addresses social risk can be used. The commenter added that dual eligibility is not a perfect measure of social risk, but is readily available so could be used as part of a short-term strategy to better incorporate social risk into clinical outcomes. The commenter provided a conceptual model for implementation.
- One commenter noted that socio-demographic factors that are outside of the control of the provider impact outcomes, and any star rating system should account for patient social risk factors in the methodology. The commenter supported peer grouping as a first step while considering approaches to account for a broad set of social risk factors.
- Two commenters noted that large hospitals, teaching hospitals, and hospitals serving a high proportion of low-income patients receive lower star ratings despite providing quality care.
 - One of the commenters urged improvements in risk adjustment approaches similar to HRRP to account for differences in hospitals and what is within the control of hospitals.
- Another commenter noted that it is misleading to the consumer to portray all hospitals as being alike, as specialty hospitals often receive five stars whereas major teaching hospitals who care for a different patient mix and offer a breadth of services do not receive the same recognition.
 - The commenter added that location and insurance coverage often influence a patient's choice of care, so it is unclear if a second peer-grouped metric would benefit consumers, and this should be examined with stakeholders before being implemented.
- One commenter stated that peer grouping could be confusing for consumers, and therefore it should only be considered an interim step to methodology improvements. The commenter recommended extending the public comment period and hosting virtual focus groups to gain more input and insights from hospitals.
- Another commenter supported peer grouping by dual eligibility status as a first step towards improved risk adjustment but noted risk adjustment itself is necessary and should be worked towards.
 - The commenter added that peer grouping could complicate interpretations and should be vetted by hospitals, physicians, patients, families, and caregivers prior to any implementation.
- Two commenters who requested direct risk adjustment noted shortcomings with the peer grouping approach used by HRRP. The commenters stated that peer grouping involves subjective choices about where to set cut-points, so hospitals in the upper end of one group and lower end of another group may have similar proportions of dual-eligible patients. Direct risk adjustment would address some of these issues, and improve precision of performance comparisons.

Of commenters who supported peer grouping, sixteen commenters supported only having one star rating by peer group, as two ratings would be confusing to the public.

- One commenter stated two ratings would be confusing to patients, would not bring value to patient decision making, and there is no clear variable for how to group hospitals into peer groups.
- Two commenters noted multiple star ratings would complicate transparency and add unnecessary burden.

Of commenters who supported peer grouping, ten commenters supported having two star ratings, one overall star rating and another based on peer grouping.

- One commenter noted that two star ratings will provide greater insight to patients and providers and stated that the *Hospital Compare* display should be clear.
- One commenter noted that if implemented, peer grouping star ratings should be supplemental to the overall star rating.
- One commenter stated that two ratings would provide the maximum amount of information to patients, who sometimes may be choosing between two different types of hospitals.
- One commenter recommended including textual and graphic display, as well as simple definitions of groupings that are understandable to the lay public, similar to benchmark data.
- One commenter recommended providing hospitals with their overall and peer grouped Star Ratings, but only publicly displaying one star rating for simplicity.

Fourteen commenters did not support peer grouping.

- One commenter stated that all hospitals should be held to the same quality standard, and a single absolute standard that all hospitals work toward is favored rather than having separate standards for different subpopulations of hospitals. The commenter suggested comparison by peer group after a single star rating was calculated, for display purposes. The commenter also added that separate star ratings may be confusing to consumers.
- Two commenters noted peer grouping would complicate transparency and add unnecessary burden.
- One commenter stated that consumers do not search according to hospital categorizations, and therefore grouping would be confusing. Patients search for hospitals that can care for their particular condition, are within a certain distance, and accept their insurance with high quality of care. The commenter added providing clarity to support consumer decision-making should be a top priority.
- One commenter stated that their users felt peer grouping was confusing to consumers, and could not agree on the best characteristics to peer group by. The commenter stated that peer grouping incentivizes hospitals to be best in a narrow group, which is opposite of the goal of the overall rating, to drive nationwide improvement in healthcare. The commenter added that quality of care should be consistent regardless of the type of, size of or distribution of services for a given hospital. The commenter instead suggested lowering the volume threshold for public reporting of measures.
- One commenter stated that although there is some concern Star Ratings disproportionately under-rate large academic hospitals, peer grouping would not provide clarity to consumers. Consumers more often look at specific geographical region for hospital comparison rather than peer groups such as academic hospitals.

- While the commenter did not support peer grouping, one commenter provided advise should CMS decide to peer group Star Ratings. They noted that exploring the impact of peer grouped star ratings would require agreement among stakeholders on the most important factors of the methodology update. If an overall star rating is retained alongside a peer grouped star rating, the commenter emphasized that differences between the different star ratings should be easily explained to consumers and hospitals should clearly understand which of the two ratings CMS values or incentivized.
- Four commenters who did not agree with peer grouping recommended instead focusing on risk adjustment.
 - One commenter felt it would be confusing to patients and instead CMS should risk adjust by bed size, safety net status, teaching status, and dual eligibility.
 - One commenter recommended improved case mix and socio-economic status risk adjustment instead.
 - One commenter suggested risk adjusting individual measures, and allowing for a filtering option by hospital characteristic on *Hospital Compare*, but keeping an overall star rating comparing all hospitals.
- One commenter added that stratification adds a level of complexity to Star Ratings, and therefore Star Ratings should be removed instead.
- One commenter noted peer grouping exploratory work is worthwhile, however there is not an obvious method for peer grouping logically or fairly. The commenter did not support peer grouping until a reasonable set of groups were defined and tested. The commenter added that performance reporting and improvement expectations should generally be independent of arbitrary groupings that could mask performance issues.
- Four commenters did not support use of peer grouping by proportion of dual-eligible patients, specifically.
 - One commenter noted it is a flawed representation of socioeconomic status since there are other factors, other than socioeconomics, that influence dual eligibility. In addition, consumers would not understand hospital categories by proportion of dual-eligible patients.
 - Two commenters noted that Medicare eligibility differs by state.
 - One of the commenters added that grouping by dual-eligible population would create state by state discrepancies.
 - The other commenter noted that this makes dual eligible status an unreliable socioeconomic indicator. The commenter added that it is also not clear that socio-economic status stratification is appropriate, as some performance indicators are related to socio-economic status while others may not be.
 - One commenter stated that the HRRP dual-eligible methodology is only appropriate for readmission measures.
 - Another commenter said dual eligibility would be a factor to control for, but not used for peer grouping, as dual eligibility affects different measures in different ways.

Seven commenters provided input regarding stakeholder concerns for the use and display of peer grouping.

- Three commenters noted that peer grouping would likely be beneficial to hospitals but confusing to consumers.
 - One commenter noted that peer grouping would be difficult to explain to consumers, who might not understand differences in types of hospitals. The commenter added that the methodology already accounts for some hospital size differences, minimum number of patients, and minimum number of measures.
 - One commenter stated that the concept may not be meaningful to consumers and the Patient & Patient Advocate Work Group recommendation should be respected.
 - One commenter noted that different star ratings would be difficult for consumers to interpret. If peer grouping is implemented, it should be clearly organized and vetted through consumer focus groups.
- Five commenters commented on the display of peer grouped Star Ratings.
 - One commenter stated that patients should be able to discern the range of services available at any particular hospital, and as currently presented patients may be unable to tell if a hospital can care for patients with more complex conditions. Peer grouped ratings should be presented within the web-based tool to assist patients in understanding the options available.
 - Another commenter stated that although peer grouping introduces complexity, consumers have the ability to recognize importance of classification, and if presented clearly, peer grouping would be valued by consumers.
 - One commenter recommended developing visual indicators for provider types (critical access, teaching) to allow consumers to easily differentiate between provider types when viewing information on *Hospital Compare*. The commenter noted it is important to provide consumers with information about different facility types. The commenter also recommended creating a separate CAH landing page.
 - One commenter suggested a simplistic approach to displaying star ratings, including a 'top hospital' within each hospital peer group.
 - One commenter supported exploring peer grouping, defined in a way that makes sense to the general public. The commenter added that it should be visually clear when searching on *Hospital Compare* which group hospitals fall into.

Ten commenters provided additional comments on peer grouping.

- One commenter stated that Star Ratings advantage specialty hospitals, hospitals with less volume and less data, compared to general acute care hospitals. The commenter added that 10% of hospitals submitting all seven domains of data are rated as one-star. The commenter therefore recommended increasing the threshold of measures required for inclusion in Star Ratings.
- One commenter noted peer grouping does not address variation at the measure or domain level, and appropriate socio-demographic status adjustment at the measure level or reporting at the service line level instead of an overall rating is needed instead.
- One commenter recommended developing a separate methodology for CAHs in alignment with CAH reporting requirements. The commenter noted that the current methodology is designed for large hospitals with high volumes of patients who report many quality measures.

CAHs do not have sufficient data and are not required to report many measures included in Star Ratings, and therefore receive star ratings based on limited data that puts them at a competitive disadvantage. The commenter added that a small change in any one measure can result in a significant change to a CAH's star rating.

- One commenter recommended a CAH-specific methodology that takes into account the disparity in services offered by different CAHs.
- Another commenter requested a solution for small rural hospitals who do not have enough cases to generate true scoring.
- Another commenter also presented analyses regarding the impact of the December 2017 Star Ratings methodology on rural hospitals. The commenter noted that analyses show significant scoring differences between rural and urban hospitals, including differences in the percentage of hospitals excluded from scoring, differences in measures reported (35 compared to 46), and differences in the mix of measures used in scoring. These differences raise questions about how effective rural hospital quality measurement is under the Star Rating methodology. The same was found for the February 2019 methodology. They recommended creating separate sets of measures for different hospital types (core set of cross cutting measures), adjusting all measures disproportionate share hospitals, Sole Community Hospitals and other facilities with larger percentages of low-income patients and uninsured patients, and creating a separate category for CAHs.
 - Based on analyses, the commenter recommended disaggregating the single rating into a more useful multi-category rating system for comparable subsets of hospitals, similar to the hospital rating system developed for US News and World Reports. The commenter recommended incorporating incorporate a separate approach for rural hospitals consistent with the National Quality Forum (NQF)Final Report on Rural-Relevant Quality Measures.
- Three commenters stated hospitals that take more transfer patients should be taken into consideration.
 - One commenter noted that patients who are transferred to academic medical centers, typically suffer from more severe, medical conditions compared to patients seen in community hospital settings.
- Another commenter stated the peer grouping decision is dependent on the goal of Star Ratings. If the goal is to inform patient health care decisions, peer grouping is not helpful as those decisions are usually based on geography. If the goal is to encourage advocacy and improvement, peer grouping by the number of measures reported would be beneficial. Peer grouping by services would be most beneficial to payers.
- Three commenters supported the incorporation of risk adjustment at the hospital level to equalize clusters.
 - One commenter suggested adjusting for hospital characteristic (bed size, safety net status, teaching status, and dual eligibility/disproportionate care etc.).

Response: Thank you for your comments and peer grouping variable suggestions. The goal of peer grouping would be to compare more similar hospitals to one another for the purposes of star ratings at the level of aggregate information summarized by the Star Ratings methodology. If peer grouping is implemented by segregating hospitals by the peer grouping variable after the summary score calculation, then the peer grouping variable that is selected would need to be relevant and meaningful

at that aggregate summary score level, or star rating level. It is also important that any peer grouping approach is consistent with the overarching goal of Star Ratings, to provide consumers and patients an easily understood summary of hospital performance. It is important to note that potential peer grouping variables should also be readily available and reliably collected on all hospitals included in Star Ratings. Public comments indicate interest and preference for numerous different peer grouping variables. Each of these variables carries different availability as well as salience to different audiences of the Star Ratings.

We previously identified and brought several potential variables (including dual eligibility, teaching status, bed size, and critical access designation) to the Star Rating TEP, Provider Leadership Work Group, and Patient & Advocate Work Group. The summary of TEP feedback is publicly available and can be found on the [CMS TEP webpage](#). The TEP and Patient & Advocate Work Group both agreed that, while potentially beneficial to hospitals, peer grouping could be confusing and unhelpful for consumers. The Provider Leadership Work Group broadly provided input that the available variables did not address their concerns of adequately accounting for differences in patient socio-economic status. While consistency with existing CMS policies and programs is a priority, the TEP and Provider Leadership Work Group both expressed concern that adjusting for dual eligibility may not make sense for quality ratings that are not tied to payment because it would create different standards of care and may not be relevant to certain quality indicators, such as safety measures, for which a patient social risk factors should not be adjusted. We will consider your input and suggestions.

Additionally, please note that many of the underlying measures are already risk-adjusted to account for patient or hospital characteristics at the individual measure level. Further adding risk adjustment in the Overall Hospital Quality Star Ratings methodology may result in measures which are, essentially, twice risk adjusted, and may mask differences in care that this rating system aims to illuminate. It is important for the current methodology to retain differences in hospital quality reported at the individual measure level in order to both accurately reflect the quality of care patients may receive as well as ensure consistency across *Hospital Compare*.

Closed-Form Solution

Several commenters supported the closed-form solution.

- Fifteen commenters supported the proposal to replace the quadrature approach with the closed-form methodology.
 - Four commenters noted the update would offer benefits and reduce barriers to hospitals and other users seeking to recreate Star Ratings using publicly released SAS Packs.
 - Two commenters suggested the update would allow timelier publication of results.
 - Four commenters reviewed the technical details and considered the closed-form methodology sensible.
 - One commenter noted the proposed closed-form solution is a well-established method.
- One commenter supported using the most efficient calculation method provided the operational impact is negligible.

- One commenter supported the update provided CMS can compare the two methods and demonstrate there is little difference.
- One commenter said quadrature is not appropriate for exponential functions like the Star Rating LVM and does not provide an accurate result; instead a closed-form solution is the only appropriate solution.

Two commenters opposed the closed-form solution, on the grounds that an alternative to LVM (such as the explicit approach) eliminates the need.

- One commenter recommended CMS prioritize other issues to improve Star Rating before implementing other changes such as closed-form solution.

Six commenters requested more information to evaluate the proposal.

- Four commenters requested additional empirical or statistical information comparing the closed-form method to quadrature before offering support.
- One commenter requested information on the costs and timeframe to implement the update.
- One commenter requested clarification on how the Constant of Proportionality (Appendix C.4, equation 4) is determined, specifically if it is fixed or hospital-specific; the commenter believes the constant should be explicitly calculated.

Several commenters weighed in on how CMS should approach similar technical modifications in the future.

- Three commenters recognized the value in general of algorithmic changes that improve speed, efficiency, and usability while producing identical results.
 - One commenter recommended that such changes be documented and made publicly available in advance of application to scoring reports to allow organizations to understand the proposed changes.
- One commenter encouraged CMS to use the most accurate and precise measurements to calculate hospital scores.

One commenter requested CMS adopt more transparency by allowing hospitals the opportunity to independently calculate Star Ratings during the preview period, prior to public release, and suggested that the closed-form solution would facilitate this.

Response: Thank you for your comments, support, and consideration.

CMS and CORE appreciate the feedback that this solution would offer benefits to hospitals and other users of the Star Rating SAS Pack. As the methodology is defined in a single SAS Pack both used to produce measure results and publicly posted for other users, the operational impact of this update is expected to be minimal. We would like to note that implementing the closed-form solution would not necessarily promote timelier publication of results, as the production timeline may be constrained by other factors, however the solution may aid troubleshooting and evaluation analyses.

Because the SAS Pack allows users to calculate the Star Rating of any individual hospital, CMS will not make the SAS Pack publicly available prior to the public reporting of results, consistent with current practices to ensure each hospital receives a confidential scoring report in advance of publication. CMS will continue to solicit stakeholder input and investigate areas of improvement for the Star Rating

methodology.

CMS believes that potentially replacing quadrature with the closed-form methodology could have multiple benefits, which include greater efficiency to assess other potential methodology changes. CMS will continue to solicit stakeholder input and investigate areas of improvement for the Star Ratings methodology.

As the methodology is defined in a single SAS Pack both used to produce the ratings and publicly posted on *Hospital Compare*, the cost of implementing this update is minimal and could be accomplished as early as the next release of Star Ratings. CMS will seek to report additional information comparing the closed-form method to the quadrature approach. CMS and CORE have performed additional internal analyses confirming the validity and reliability of this method compared to the quadrature approach. CMS intends to publish additional information reflecting this in the future.

CMS will continue to investigate areas of improvement for the Star Ratings methodology while soliciting stakeholder input and alerting stakeholders to any prospective changes.

Long-Term Future Considerations

Explicit Approach

Ten commenters supported the LVM with modification, rather than changing the methodology to an explicit approach.

- Four comments supported changes measure weighting, as suggested with an explicit approach, but only if LVM is maintained.
 - One commenter specifically recommended fixed weights with the LVM.
 - Another commenter notes that the LVM model is more flexible to the addition of new measures and difficult to *game* the system. The commenter recommends using equal weights for the measures and exploring empirical Bayes estimators. Additionally, the commenter notes, the explicit approach is likely not data driven, the process having to be repeated when any measures are added or dropped. It is recommended that a hybrid process is considered where LVM is used to develop a range of weights under various assumptions and those weights are brought to stakeholders for discussion.
- While the commenter supported the current LVM approach, one commenter noted confusion regarding the precise impact and significance of likelihood weights in the LVM approach. They suggest a measure that impacts three times as many patients as another should be weighted at least three times as much.
- One commenter notes that explicit approach leads to regulatory inclusion of measures and thus, the current methodology should be enhanced and implemented.
- One commenter supported the LVM but also believed exploration is warranted to ensure stakeholder responsiveness.

Four commenters expressed concerns with several aspects of the explicit approach.

- One commenter noted the explicit approach would be an artificial way of weighting measures.

- Another commenter recommends CMS examine current measures to evaluate which should be altered or retired to allow more fairness of scoring across measures. They note that an explicit approach can single out measures for improvement on an annual basis.
- One commenter suggested using electronic health record (EHR) data to ensure sustainability of the explicit approach.
- One commenter suggested continued involvement of stakeholders to ensure best understanding of data analytics.

Forty commenters supported the explicit approach methodology.

- Twenty commenters provided comments on an explicit approach with a simple average.
 - Fifteen commenters noted that replacing the LVM methodology with the explicit approach would address the lack of understanding, consistency, transparency, reliability, and accuracy.
 - One commenter recommended not using the simple averaging approach when implementing the explicit approach.
 - One commenter noted that using the current LVM does not allow hospitals to develop an understanding of how they can improve patient care, does not improve quality improvement efforts, and increases administrative costs. Therefore, the commenter recommends retiring the LVM model for an explicit approach. The same commenter notes that the current groups of measures are arbitrary and supports their statement from articles by the Association of American Medical Colleges and Modern Healthcare.
 - Another commenter explains that the current model does not reflect accurate hospital quality efforts and the statistical methodology is very challenging to understand, even for educated consumers. This commenter also notes that Star Ratings is currently being used by private sector payers and this is hindering hospital outlook as the star ranking cannot be predicted and although the hospital may be improving, this is not apparent. The commenter recommends the systematic surveying of patients to identify the aspects of quality that is most important to this population and then use this data to implement weights.
 - One commenter notes that the explicit approach aligns with the approach used for the Medicare Advantage and Part D Star Ratings, as well as other hospital quality report cards and so they support the explicit approach.
 - Another commenter recommended CMS to implement an explicit model to improve transparency and accuracy of each measure contribution to performance.
 - One commenter supported the use of an explicit approach that provides rationale for dominant measure loadings.
- Thirteen commenters expressed support for select advantages of an explicit approach.
 - Five commenters noted an explicit approach would be easier to understand for stakeholders.
 - Another commenter noted that the explicit approach is predictable for hospitals and assesses all hospitals in the same manner. They explain that the weighting can be

- dependent on strength of evidence, impact on patients, and/or hospital opportunity for improvement.
 - Five commenters noted that predetermined weightings would allow hospitals to focus on key quality improvement and performance initiatives.
 - One commenter noted the current methodology makes it impossible for hospitals to achieve a 5-star rating and a simplified approach would allow for understanding of actual performance for both hospitals and consumers.
 - Two commenters noted that the explicit approach allows for transparency and reliability.
 - One commenter noted this approach would still require education for consumers.
- Fifteen commenters commented on the process for determining measure weights.
 - Four commenters recommended an interdisciplinary group develop the weighting system where they are encouraged to explain their rationale clearly.
 - One commenter suggested CMS consider focus groups to determine clinical significance weights to measures to keep this approach sustainable.
 - Three commenters recommended a more evenly distributed weighting system.
 - One commenter noted although they support explicit scoring, they acknowledge challenges around determining the weights for each measure.
 - Another commenter recommended tailoring weights to the measure set but also balancing across measure groups to avoid too much emphasis or too less emphasis on one measure.
 - Another commenter recommended using empirically based approaches (such as surveying patients on measure importance) to make measure weight decisions.
 - Another commenter noted that in addition to fixed weights, final scores should be risk adjusted.
 - One commenter agreed with pre-defined measure weighting.
 - One commenter supported a methodology that would allow for predetermined measure weighting with more balance across measures.
 - Another commenter supported prespecified differing weights, and to use a methodology that enables accuracy, reliability, and the ability for patients to use a rating system that truly reflects quality.

Seven commenters urged for alignment between CMS programs, as it relates to an explicit approach and measure weights.

- Five commenters asserted there should be consistency between all CMS reporting programs including: Overall Hospital Quality Star Rating, Hospital Value Based Purchasing (HVPB), HRRP, and Hospital Acquired Conditions Reduction program (HACRP).
 - One commenter encourages CMS to model Star Ratings after the HVPB program which allows low-performers to rise rather than stay at the stagnate at the bottom. Additionally, this commenter also encourages CMS to consider a harm-based weighting similar to the Agency for Healthcare Research and Quality's (AHRQ) PSI-90 component weighting.

- One commenter specifically noted that the defined set of measure weights in the explicit approach methodology should be similar to HVBP and HACRP programs. They recommended distributing weight from missing to measures to other measures within the group.
- One commenter noted the same measures are used in hospital pay for performance programs.

Two commenters provided additional comments on incorporating the explicit approach.

- One commenter stated that they do not support the current LVM model or the explicit approach. The commenter recommends CMS to explore more data driven solutions.
- One commenter stated current complexity of the methodology obscures true hospital performance.

Response: Thank you for your comments. The original objective of Star Ratings was to provide consumers with a summary of the many quality measures on *Hospital Compare*. However, we realize that star ratings are realistically used by different audiences, including providers to improve quality of care, and the methodology may be difficult to interpret or communicate in certain cases.

The decision to implement Star Ratings with the LVM approach was based on a series of research, analyses, and stakeholder engagement activities. LVM generates a single measure group score as a single quality trait through empiric, objective methods of combining measure information while also accounting for missing information and the relationship between measures within a group. In other words, the statistical modeling used data to determine measure loadings that reflect correlations between measures and are the same for every hospital, despite differences in hospital measure reporting. Despite the advantages, we acknowledge that measure loadings can shift between periods and it may not be apparent to stakeholders what caused the change, leading to confusion and inability for hospitals to explain changes in subsequent updates to scores and star ratings.

An explicit approach would be easier to understand and explain. However, a simple explicit approach may not account for measure precision or differences in measure reporting. In the simplest form of an explicit approach, each measure would have equal weighting within a measure group, regardless of volume, and individual measure weighting would be different for each hospital. For example, if hospital A has three measures within mortality but hospital B has all seven measures, hospital A's three measures will each be weighted 33% but hospital B's seven measures will each be weighted 14%. Furthermore, given the evolution of hospital measure reporting and changes to the overall measures available on *Hospital Compare* over time, the consistency of emphasis on individual measures for each hospital or overall may not be preserved.

We will consider your comments and feedback.

Alternatives to Clustering

Four commenters supported the current k-means clustering approach to Star Ratings.

- One commenter noted any alternatives to k-means clustering would likely have minimal effect on period-to-period predictability as hospitals' relative summary scores are driven by the LVM.
- Two commenters stated that although the current approach is not predictable or easily reproducible for stakeholders, it is methodologically sound.

Five commenters noted current k-means clustering is opaque to stakeholders and urged for more transparency in the Star Ratings methodology.

- Two commenters requested a simpler system for hospitals to be able to replicate and manipulate for their own analyses.
- One commenter requested publicly available cluster analyses and assessment statistics.

Nine commenters were not supportive of the current k-means clustering approach as individual hospital Star Ratings are dependent on the summary scores of other hospitals.

- Two commenters noted this method unfairly forces hospitals into five groups, despite small variance in hospital summary scores.
- One commenter stated k-means clustering inaccurately captures individual hospital performance and introduces period-to-period uncertainty for borderline hospitals. They urged for hospitals to be rated solely on their own performance.
- One commenter asserted clustering would be unnecessary if hospital star ratings had a flat or normal distribution.
- One commenter suggested using decimal ratings between 0.0-5.0 and scaling the final normal distribution of scores to the mean, equivalent to 3.0 stars.
- One commenter noted the current k-means clustering approach is suboptimal as the underlying data is “messy”. They urged for the implementation of performance standards to promote period-to-period and in-hospital consistency, and suggested gaining stakeholder feedback on k-means clustering.
- One commenter noted k-means clustering assumes the validity of a forced bell-curve solution and urged a more linear approach.

Twelve commenters expressed concerns about the current k-means clustering approach.

- Six commenters noted the use of k-means clustering decreases period-to-period predictability, hindering hospital improvement efforts.
- One commenter noted the Star Ratings is not tied to performance metrics and there are no standards within Star Ratings which are independent of other CMS programs.
- One commenter added k-means clustering is typically used on multidimensional measures, whereas the Star Ratings methodology is based on a single dimensional measure.
- One commenter suggested using a clustering approach similar to that of Medicare Advantage and Part D Star Ratings.
- One commenter proposed including historical hospital performance in order to smooth summary scores to decrease period-to-period changes.
- One commenter stated that results from their analyses indicate the current clustering approach has low reliability for 5-star and 2-star hospitals.
- One commenter supported using empirical methods for determining Star Ratings cut-points. They noted the benefit of the k-means clustering is it allows for uneven group sizes. However, the commenter suggested CMS explore multiple clustering factors as it sheds light on relationships between different factors and hospital quality. The commenter also noted the selected method should produce valid, understandable and consistent results.

Thirteen commenters suggested hospitals be clustered by policy-determined cut-points.

- One commenter added these cut-offs would need to be revised as all hospitals improved over time.
- Five commenters proposed assigning stars based on performance percentiles of policy-determined cut-offs, or by normalizing the summary scores.
- One commenter noted performance percentiles would be more consistent with HACRP.
- One commenter suggested incorporating policy-determined performance categories into the calculation of each measure group.
- One commenter stated that set cut-points would better support quality improvement efforts as hospitals will have clear benchmarks to work towards.
- One commenter suggested the Mean Shift Clustering method as it conveys the underlying probability of score distributions.

Response: Thank you for your comments.

The current Overall Hospital Quality Star Ratings methodology report is publicly available on www.qualitynet.org, in addition to the Quarterly Updates and Specifications report, which documents the underlying methodology used to calculate the Star Ratings for each refresh on *Hospital Compare*. Additionally, the data for each reporting period, the SAS pack, and supplemental documentation guide has been made publicly available for each public release of Star Ratings since the first release in July 2016 to ensure complete transparency of the Star Ratings methodology and calculation.

The current k-means clustering approach, which was vetted by stakeholders in the development of the original methodology, intentionally compares the summary scores of hospitals in order to ensure that there are five groups of summary scores that are more similar within groups and more different between groups. Validity analyses consistently demonstrate that the Star Ratings are effectively assigning hospital star ratings in that hospitals with higher star ratings perform better on the underlying measures and vice versa.

Although changes in star ratings are the results of several factors, including changes in hospital performance and added, removed, or updated measures. Previous analyses have revealed that there are hospitals with summary scores at the borders of the k-means clusters and those hospitals may experience slight shifts in star ratings between reporting periods. However, we expect this issue would persist with alternative approaches since almost every approach requires cutoffs and there will always be hospitals at the borders of those cutoffs.

CMS will consider your feedback and continue to evaluate the k-means clustering methodology and its impact on hospital star rating assignments.

The rationale for the current k-means clustering approach is to group hospitals with similar summary scores such that any given hospital's performance is most similar to other hospitals within that group. This approach was vetted by experts in the original development of the Star Ratings methodology. Further considerations would need to be made before implementing any potential set cut-points or using percentiles to categorize hospitals into the five groups.

Incorporation of Improvement

Eleven commenters supported the incorporation of performance improvement within the Star Ratings methodology.

- Five commenters supported the incorporation of performance improvement because it would incentivize and reflect quality improvement efforts.
- One commenter felt giving hospitals credit for improvement could be important for smaller communities with less hospitals in the area.
- One commenter felt that improvement should be added if the intent of Star Ratings is to provide information to health systems and consumers about quality.
- One commenter said they support an improvement methodology that uses a rolling average in the star ratings calculation.
- One commenter felt that improvement should be incorporated into Star Ratings, but felt there were concerns with the definition of improvement and how it would be incorporated, including the idea that this might punish already high performing hospitals.
- One commenter supported the idea of incorporating improvement saying that poor performance should be for hospitals that have a low score and are also not showing improvement.

Sixteen commenters did not support the incorporation of performance improvement within the Star Ratings methodology.

- Six commenters did not support incorporation of improvement given that the goal of Star Ratings was to rate hospitals for consumers and the star ratings would no longer be a summary of quality if improvement was added.
 - Two commenters specifically noted that incorporation of improvement would not align with the Patient & Advocate Work Group and general consumer requests for a simple view of current quality data.
 - One commenter said incorporating improvement would be deceptive and called it “incorporating bonus points.”
- One commenter did not support including improvement as it would be confusing for the public.
- Two commenters said that incorporating improvement would add complexity.
- Four commenters expressed concern that the approach may disadvantage high performing hospitals with four- or five-star ratings.
 - One commenter added they would want further details on the approach as they would not want improvement to inflate scores beyond hospitals who are consistently high performers. They also requested more details on “absolute improvement”.
 - One commenter asked what would be the benefit for hospitals that stay consistently at a five- or four-star rating, stating that a hospital that improves from a lower star rating may benefit more than a hospital with consistently high star ratings.
 - One commenter pointed out that incorporation of improvement is currently used in the HVBP program and they feel that this approach only incentivizes poor performers.

- Five commenters suggested reporting previous ratings, HVBP improvement, or some indication of improvement on *Hospital Compare*, instead of incorporating improvement into the Star Ratings methodology.
- Four commenters did not support incorporating previous periods of data in the calculation as most current data should be used.
 - One commenter pointed out that some of the individual measure data already reflects older performance from one to two years ago.
 - One commenter did not support a method that would “smooth” year to year changes and suggested placing more weight on groups with more current data.
- One commenter did not support the inclusions of improvement given the variability of the LVM.
- One commenter did not support including improvement as the overlapping reporting timeframes might not provide a clear picture of improvement.
- One commenter did not support adding in improvement as a change in score between reporting periods would be reflected in a better or worse refreshed rating.

Thirteen commenters provided other comments on the topic of incorporating performance improvement within the star ratings methodology.

- Commenters provided suggestions for implementing the incorporation of improvement within the Star Ratings methodology.
 - Seven commenters mentioned using an approach similar to HVBP.
 - One of these commenters recommend that greater weight be given to absolute score to avoid punishing the highest performing hospitals.
 - Another commenter recommended that CMS choose a methodology that does not penalize consistently high performing hospitals. They also noted that incorporating absolute score would be transparency for patients, but the risk is that higher ranked hospital could be perceived negatively from one year to the next.
 - One commenter suggested the Star Rating should have a top decile and top quartile scoring process for achievement thresholds.
 - One commenter recommended using a hold-harmless policy if an improvement was included in Star Ratings so that high performer hospitals are not penalized.
 - One commenter recommended incorporating measures of year-to-year differences. The commenter added they felt that consumers would be more interested in actual performance of hospitals compared to improvement.

Response: Thank you for your comments. CMS has not received much support from stakeholders on incorporation of improvement in the past. Stakeholders have expressed support for the Star Ratings to represent the most current data for the underlying measures. Also, the primary purpose of Star Ratings is to summarize quality information for consumers, not to facilitate or direct quality improvement activities for which individual measures provide higher fidelity. Additionally, we received feedback suggesting the use of complimentary icons, such as an up arrow, if improvement is incorporated at all, rather than changing the ratings themselves.

Nineteen commenters supported user-customized Star Ratings.

- Two commenters emphasized that the intent of Star Ratings is to provide consumers with quality information; they should be able to adjust the information to reflect their needs.
- One commenter stated that, in the 21st century, patients deserve high-quality methods that are both personalized and precise.
- One commenter added that user-customization would be based on who is using the site.
- One commenter stated user-customization is worth exploring and suggested an overall star rating does not serve consumers well.
- One commenter supported a sophisticated consumer tool but noted it may not be used frequently, and much of the data is already available in other avenues on *Hospital Compare*.
- Two commenters suggested a customization similar to that of a recently published tool would be helpful.
 - One commenter noted this tool would require training for the consumers. The commenter added the rollout of such a tool would require multiple weights based on consumer profiles or backgrounds.
- One commenter stated this tool aligns with the direction of precision in medicine. The commenter suggested CMS conduct a pilot test of the tool.
- One commenter supported user-customized star ratings but expressed concern that the concept may be premature, noting weights consumers place on measures may not reflect true quality of care provided.
- One commenter suggested additional analyses be conducted to ensure customized ratings are reliable and valid.
- One commenter supported user-customized star ratings but proposed allowing the consumer to select measures they find relevant to themselves.

Sixty-two commenters did not support the user-customized star ratings.

- Five commenters proposed a star rating for each measure group, rather than same as, below, or above the national average, in place of or in addition to the overall star rating.
 - One commenter added that a diverse group of stakeholders should be involved in the development and testing of this proposed change.
 - One commenter explained more work needs to be done to ensure consumers understand the results.
- Three commenters noted this tool should not be incorporated into the Star Rating methodology; if developed, it should be a separate tool overlying the same data set.
- One commenter noted that user-customized star ratings would not provide hospitals with feedback for improvement.
- Forty-one commenters flagged that user-customization would be too confusing to consumers.
 - Twenty-nine commenters additionally expressed concern over whether consumers understand measure groups or what is measured within them, noting this model assumes general consumers have working clinical and statistical knowledge.

- One commenter proposed incorporating common clinical conditions with sets of weights appropriate to the conditions. The commenter also suggested separating hospitals by clinical conditions as proposed by the American Hospital Association (AHA).
- One commenter added that the measure group level information was already available on the *Hospital Compare* site.
- One commenter said a user-customized tool might lead to confusion because small differences in absolute risk might be over-weighted (for example, some procedures are much riskier than others but consumers may not grasp the differences in absolute risk of different measures). The commenter noted that weighting of different metrics is still debated by academics and clinicians and said it is potentially unsafe to push this responsibility onto the consumer.
- Two commenters explained there are many factors that patients consider when selecting a hospital for their care, and that a hospital's star rating might not capture that hospital's performance in an area of care critical to the consumer.
- Two commenters proposed displaying star ratings for each measure group.
 - One commenter recommended this in addition to the overall star rating. The commenter also suggested additional filtering features on *Hospital Compare* to allow users to filter individual measure group stars.
- Five commenters noted not all conditions have measures in place, therefore not providing consumers with metrics that define the care they may be looking to evaluate.
 - One commenter flagged this could result in inaccurate Star Ratings.
- One commenter stated they do not support the tool as it would be difficult for hospitals to manage.
- One commenter noted the proposed tool is too complex.
- One commenter suggested more analyses be conducted to better understand consumer interest and understanding of the proposed tool.
- One commenter instead suggested a framework similar to New York State Department of Health's consumer guides for health plans, where each domain has its own star rating.
- Four comments did not support the development of a user-customized tool prior to addressing other existing concerns about the methodology.
 - One commenter recommended revising the methodology to create more consistent ratings prior to allowing consumers to customize star ratings.
 - One commenter noted that user-customization does not address consumer concerns regarding predictability and transparency of measure group weighting.

Response: Thank you for your comments.

The original objective of Star Ratings was to provide consumers with a summary of the multiple quality measures on *Hospital Compare*, and user-customized star ratings would be consistent with the original intent. However, we realize that Star Ratings are realistically used by different audiences, including providers to improve quality of care, and user-customized star ratings may not be as helpful for hospitals to target quality improvement efforts. CMS recognizes the complexity in understanding consumer design and content preferences for customized tools such as this application and will engage a broad array of consumers in iterative testing if this project proceeds. We will consider your comments and feedback.

Additional Comments on the Overall Star Ratings

Overall Project

One commenter acknowledged that no rating system is perfect and that the underlying quality measures and information are still maturing.

Eleven commenters agreed with the objectives behind the Star Ratings or the importance and need for an overall quality rating.

- However, one commenter noted that categorizing, advertising, and consumer use of quality ratings should be studied further.
- Two commenters specifically agreed with the objective to make the information on *Hospital Compare* easier for consumers to use and understand to assist in their decision-making.
- Another commenter advocated for changes to the system to support the goal of providing patients with a clear, simple and objective mechanism for identifying top performing hospitals.
- One commenter noted that with methodological revisions, the Star Ratings have great potential to aid consumer choice and provide hospitals with meaningful quality metrics to promote improvements in patient care.
- Another commenter noted the value of having an easily accessible tool for consumers, but felt extensive methodology changes and simplification was necessary.
- One commenter noted that despite the flaws in the Star Rating system, they are committed to improving performance and look forward to improving the methodology to better inform patients on quality of care.
- One commenter supported the overall drive towards value-based care, and availability of information to consumers to make educated decisions.
- One commenter agreed with the need to evaluate quality across hospitals and acknowledges the challenges of creating a composite quality rating. The commenter also noted that there will always be concerns from hospitals that do not perform well and these concerns shouldn't invalidate quality efforts or dissuade CMS from continuing to publish the ratings even as methodology updates are being considered.
- One commenter supports the development of Star Ratings, but expressed concern over its validity and usefulness.

One commenter urged CMS to keep and continue to improve the Star Ratings as well as CMS payment programs since these efforts drive true hospital quality improvement in a way that other compare sites alone could not.

Thirty-seven commenters recommended that CMS either remove entirely or suspend Star Ratings until the methodology is refined. While almost all commenters supported CMS's effort to provide additional transparency on hospital quality, they shared the following concerns regarding the Star Ratings.

- Two commenters recommended suspending Star Ratings until methodology updates are made, CMS engages an independent auditor to review the methodology updates, CMS removes the Efficient Use of Medical Imaging measure group, the methodology is transparent

and reproducible by hospitals, and stakeholders and the public are educated on the methodology.

- One commenter noted insufficient information was provided by CMS for the stakeholder to provide input on the impact the proposed changes will have. The commenter suggested CMS suspend the release of Star Ratings until CMS provides more detailed information, and conduct another public input period where CORE releases its research database and SAS pack for hospitals to replicate analyses.
- One commenter stated until the methodology is improved, it is difficult for consumers or hospitals to have confidence in the ratings.
- One commenter noted disappointment that Star Ratings continue to be available while the methodology is being modified. They suggested suspending the ratings until there is time for stakeholders to understand the current methods and proposed changes.
- One commenter stated the Star Rating should be suspended until a sound methodology that adjusts for socioeconomic status and peer grouping that is driven by actual performance is developed. The commenter stated that continuing to display the current ratings adds burden and cost to hospitals and is misleading to consumers.
- One commenter elaborated that the delay in release of Star Ratings, hospital consensus that the ratings have no face value, as well as the lengthy technical report posted for public comment indicate that CMS should remove the current ratings and focus on improving the methodology. The commenter further commented that explaining the current Star Ratings is taxing on hospital leadership and confusing to patients, especially when hospital resources should be focused on improving quality of care and educating consumers.
- Another commenter stated that there are unintended consequences of retaining the current Star Ratings. In addition, payers and other stakeholders are using Star Ratings to determine reimbursement.
- An additional commenter suggested that, instead of continuing to publish the Star Ratings, CMS improve the individual measure information on *Hospital Compare* so that it is easier for consumers to use and understand. If CMS chooses to continue the Star Ratings, the commenter suggests suspending the Star Ratings until the methodology is updated.
- Two commenters recommend that CMS examine feedback received and modify the methodology for more accurate ratings before the next release of Star Ratings.
- Another commenter recommended suspending Star Ratings until stakeholders come to agreement on risk adjustment, stratified reporting, and a clearer methodology that does not penalize safety-net organizations.
- One commenter recommended suspending the Star Rating and recommended investigating user-customization based on performance thresholds, removing the LVM for the Safety of Care measure group, emphasizing the importance of current quality, avoiding clustering techniques, and testing the robustness of modeling decisions.
- Two other commenters recommended holding publication of ratings until methodology considerations were addressed to avoid confusing patients, and CMS should allow time to fully vet the proposal and allow stakeholders time to fully understand changes.
- Two commenters suggested CMS thoroughly investigate methodology updates to avoid disadvantaging any hospital types.

- One commenter recommended removing the Star Ratings until all proposed changes are fully vetted with stakeholders to ensure patients have meaningful and accurate quality information.
- One commenter recommended suspending Star Ratings and addressing transparency, clear cut-points and targets, and accurate data reflecting quality of care provided.
- One commenter explained that CMS does not provide educational materials for consumers to interpret the Star Rating and the Star Rating as a summary does not reflect socioeconomic factors.
- One commenter noted that the ratings cause confusion and risk to the health and well-being of patients and should be suspended until improvements are made.
- One commenter recommended that CMS convene a stakeholder work group to provide input on the methodology. They further suggested that CMS review a study published by Rush University that outlines challenges associated with star rating performance, including outlier readmissions, readmission adjustment, SES adjustment, and variability in ratings due to LVM.
- One commenter recommended eliminating Star Ratings permanently and instead focusing on individual categories related to care.
- Another commenter recommended removing Star Ratings from public view, even with proposed methodology changes.
- One commenter suggested eliminating the Star Rating for a year or more until there are no significant shifts in ratings from one period to another.
- One commenter supported suspending the release until validity and reliability are improved.

Twenty-eight commenters expressed concern that Star Ratings are not serving the original purpose as a reliable, transparent quality tool for consumer use.

- One commenter stated that Star Ratings, at this time, do not achieve the aim of a transparent measure of quality that is easy to understand by consumers and healthcare quality leaders.
- Another commenter stated the Star Ratings fail to provide patients with an accurate representation of quality for decision making, and the lack of transparency of the methodology does not allow providers to understand scores.
- Two comments stated the current ratings are inaccurate and misleading to patients and consumers.
 - One of the commenters stated that they support transparent, valid and meaningful information for patients to make care decisions, but the current ratings are harmful to patients forming incorrect conclusions about hospitals.
- One commenter expressed concern that the ratings cause confusion, rather than assist consumers, as they seek more information about healthcare quality. The commenter supported the decision to not publicly report the July 2018 Star Rating results.
- One commenter stated it is counterproductive to release ratings that misrepresent quality of care provided by hospitals. The commenter added the methodology does not achieve the goal of providing transparent quality information to consumers and healthcare leaders.
- One commenter stated that Star Ratings do not serve as an effective tool for displaying individual hospital quality or comparing hospitals, regardless of care settings.
- One commenter stated they do not believe the ratings provide their community with accurate information regarding the quality of health care services available to them.

- One commenter noted concerns with the Star Rating methodology and potential updates outlined in the public comment document.
- An additional commenter stated that the Star Ratings do not reflect differences in function and patient populations for academic medical centers and are not reliable enough for patients to make informed decisions about where to seek care.
- One commenter stated CMS should not post data that is more likely to confuse patients rather than assist them, and that provides no direction for improvement strategies. The commenter stated cohesive, meaningful measurement programs are necessary to allow for understanding.
- One commenter stated that the Star Ratings may cause more harm than benefit and should be revised or eliminated. The commenter further explained that the individual measures were not developed to be combined into an overall rating and do not result in reliable or equitable comparisons.
- One commenter stated they support statistically sound methods, but questioned whether the ranking system reflects consumer priorities.
- One commenter stated that the current Star Ratings do not align with patient priorities, do not accurately reflect quality, and are unpredictable. The commenter further questioned whether patients find the Star Ratings meaningful and useful. They also stated that the Star Ratings do not align with any CMS initiatives, which do not usually align with patient priorities for network coverage, travel time, and services provided. The commenter encouraged CMS to review the usefulness of the Star Ratings before evaluating technical methodology updates.
- Two commenters expressed there are not enough patient-reported outcome measures that reflect the needs of patients, so Star Ratings would not be useful in directing a patient to the best facility for their care.
- One commenter stated that because the rating combines multiple dimensions of quality, it may not provide information that is most important for a patient's situation. The commenter recommended that CMS explore other comparative approaches, and that the rating only provides a summary of a few discrete processes of questionable representation of quality and outcomes.
- One commenter noted that patients typically choose a hospital based on location and their condition and need multifaceted information to make their decision. Therefore, an overall star rating is likely not meaningful or useful.
- One commenter stated the ratings should provide transparency, continuity, and reliability and allow providers and patients opportunity to understand the measures and calculations that make up the ratings. The measures and calculations behind the Star Ratings should improve care and outcomes.
- One commenter stated they have analyzed each Star Rating update to determine if the methodology has served as a resource for patients with statistical objectivity, and expressed concerns with the unstable results that have been shared. The commenter added that consumers rely on statistical modeling to provide objective assessments.
- Two commenters did not favor the comparative approach to Star Ratings.
 - One commenter stated that the Star Ratings do not provide fair or equitable comparisons of hospitals.
- One commenter stated that the Star Ratings do not correlate with individual measure results. In particular, the measure-level performance categories of better, no different, or worse than

national average do not align with the ultimate star rating. The commenter expressed concern that the Star Ratings do not reflect measure performance and consumers will not investigate quality information beyond the Star Ratings, providing a potentially inaccurate summary of quality.

- One commenter stated that the goal of a meaningful, simplified snapshot of hospital quality is compromised by an inaccurate and misleading comparative methodology.

Sixteen commenters noted that the Star Ratings are an oversimplification of quality information.

- Three commenters expressed concern about potential consequences for patients making decisions using an oversimplified representation of hospital quality.
 - Two commenters stated that the measures were not developed with the intent to be displayed as part of a composite.
 - One commenter noted that this is particularly true for CAHs where there is a great disparity in services offered.
- One commenter questioned whether summarizing meaningful variation in a broad set of measures into a single rating in a valid and fair way is achievable.
- One commenter stated a single rating that combines diverse measures oversimplifies complex clinical factors and does not provide fair and meaningful information. The commenter added that this is particularly true for teaching hospitals that care for a more sick and complex population.
- One commenter also noted that the Star Ratings are based on measures that fail to adjust for complex patient medical conditions and sociodemographic factors.
- One commenter noted that an overall composite rating misrepresents complexity of care for a large volume of diverse patients with multiple, complex comorbidities.
- Another commenter stated that a single overall rating may not provide patients with meaningful information for the specific care they need.
- One commenter stated the notion of a single composite score is flawed and over simplifies complex data that is not representative of all cases.
- An additional commenter stated a single composite that combines measures oversimplifies many complex factors and may not represent overall quality and important outcomes. The commenter added that although they support transparent information provided to patients, it must be displayed meaningfully and reflect importance for the patient's situation.
- Two commenters stated that measures cover a wide variety of clinical areas, but only a handful of measures ultimately impact the star rating regardless of the number of measures the hospital reports.
- One commenter noted that reducing the many services and aspects that contribute to quality of care to a single rating is drastically simplified.
- Three commenters stated that CMS should ensure the Star Ratings do not oversimplify a complex and individualized decision, while potentially exacerbating disparities in care.
 - Two of the commenters added that the rating must reflect cross cutting measures that affect all patients, that are relevant to their care choices. The 57 measures included in Star Ratings do not allow for a single rating that reflects all aspects of care. The

commenter added that because each patient's circumstances differ, so will the measures that matter to them.

Thirteen commenters supported transparency in healthcare and providing clear, meaningful quality information to the public to facilitate healthcare decisions but expressed the following concerns.

- One commenter noted that this is challenging when there are complex, multifaceted differences of quality across hospitals serving different population groups and offering difference services.
 - The commenter noted that, while they perform well on Star Ratings, they believe the ratings require modification before it can be a true and effective tool for patients and hospitals
- Another commenter supported providing patients and consumers with clear and meaningful quality information and urged CMS to improve the methodology which may lead to inaccurate quality comparisons.
- One commenter commended CMS's resolve to improve the usability, accessibility and interpretability of *Hospital Compare* for patients and consumers.
- Two other commenters noted concern for the usefulness of Star Ratings due to the methodology approach.
- One commenter stated that the intent of Star Ratings should be reconsidered, as the goal to include as many measures as possible does not align with the Meaningful Measures Initiative. The commenter added that measures used to direct improvements do not necessarily translate into useful information to direct patient choice.
- Two commenters urged for a methodology that reflects aspects of care most relevant to the consumer.
 - One commenter stated the existing methodology may cause confusion and poor decision-making, especially when a hospital's rating does not align with performance on the specific condition or procedure for which the patient is seeking care. The commenter recommended providing a subset of measures in place or in addition to a composite score, which would also help drive hospital quality improvement efforts.
 - One commenter added the methodology should be accurate through meaningful measures and correctly executed statistical modeling, stable, and predictability as well as reflect measure performance, a balanced assessment of multiple measures, and account for potential clinical and social risk factor biases.
- Another commenter stated the Star Ratings provide misleading and inaccurate information, although transparency in cost and quality in health care is important.
- One commenter questioned the value of a hospital star ratings system since the existing measures are many and uncorrelated and star ratings based on these measures would have little to no predictive power for a particular situation. For example, a patient seeking quality information on elective spine surgery would get more representative information from related individual measures. Furthermore, some measures are weighted more than other measures, giving undue influence to particular sets of measures, and measures are placed into categories, even if differences between categories are not clinically meaningful. Instead of star ratings, patients could review case or procedure volume at a given hospital to determine whether the hospital is capable of providing the services they need. Poor quality of

care may be a consequence of hospital staff trying to provide care outside of their scope of capability.

Ten commenters stated that the Star Ratings do not provide actionable information for quality improvement.

- Three commenters noted it is difficult to focus on particular measure, when the loadings change each reporting period.
- One of the commenters added that because of the statistically complex methodology, hospitals cannot predict their performance or predict measure performance impact on their rating. The commenter noted although the stated goal is for use by patients; however, hospitals use publicly reported information to drive quality improvement, and a more predictable star rating would allow for better benchmarking by hospitals.
- One commenter stated the methodology has led to inconsistencies and does not allow hospitals to predict their scores.
- Another commenter noted hospitals cannot be held accountable without performance targets, and clear cut-points known ahead of time by hospitals, so hospitals understand quality performance goals.
- One commenter noted that providers expect reliable and consistent measurement for quality improvement, and the current Star Rating do not meet this goal to provide accurate hospital comparisons in heterogeneous hospitals.

One commenter stated the Star Ratings are a CMS invention not mandated, required, or guided by regulations.

One commenter suggested individual states publish state-wide star rating systems independent of the federal government.

Response: Thank you for your support and feedback. The primary objective of the Overall Star Ratings is to summarize the existing quality information on *Hospital Compare*, which currently publicly reports on over 100 measures for over 4,000 hospitals, in a way that is useful and easy to interpret for patients and consumers. Many patient advocacy groups, government entities, and purchasers have requested clearer display options of *Hospital Compare* for consumers and have expressed support for the launch of the Overall Star Ratings. CMS' support contractor, CORE, convened and worked closely with a Patient & Advocate Work group to guide development and solicit input on the need for and the usefulness and meaningfulness of the Overall Star Rating. Consistent with other CMS Star Ratings, the summary assigns each hospital between one and five stars. In addition, performance categories of above, below, or same as the national average are provided for each of the seven measure groups.

CMS is committed to providing the public with transparent hospital quality information in easy to understand and familiar ways, such as Star Rating metrics, as they navigate the healthcare system and make difficult decisions during vulnerable times. While quality improvement was not part of the original objectives of the Star Rating, CMS acknowledges that the Star Ratings may be used by multiple stakeholders, including providers. As such, the Star Ratings are always reported alongside individual measures that are better suited towards hospital quality improvement initiatives. CMS will continue to engage stakeholders as well as monitor the usability and methods of the Star Ratings.

Overall Methodology

Nine commenters were concerned about the complexity of the Star Ratings methodology.

- One commenter used the closed form solution as an example of unnecessary complexity within the Star Ratings methodology.
- Another commenter stated that the complexity of the methodology prevents providers from being able to communicate the meaning of Star Ratings in a thoughtful and clinical manner. The commenter further elaborated that patients may be frightened or confused by the complexity of Star Ratings.
- Two commenters stated the complex statistical techniques used lack transparency and create uncertainty. The commenter added the flawed methodology drives ratings, not hospital performance.
- An additional commenter stated the methodology should be simplified so all stakeholders, including the public, can understand.
- One commenter noted that, as a hospital association, they encounter challenges in understanding and explaining the methodology to consumers and clinicians.
- Another commenter supported a simpler approach that fits the dimensionality of the targeted measures, to increase transparency to consumers and providers.

Three commenters suggested alternative methodologies to the Star Ratings.

- Based on feedback from a Patient Family Advisory Council, one commenter suggested adding an audience ratings system, similar to Rotten Tomatoes, in addition to the current methodology.
- Based on feedback from a Patient Family Advisory Council, one commenter suggested the use of domain star ratings, in which hospitals receive separate star ratings for each measure group.

One commenter states the methodology should be based on appropriately executed statistics, and differences in ratings between hospitals should be substantiated by differences in underlying measure performance.

One commenter specifically noted Rush University's research on the methodology and noted it as evidence for reform of the Star Ratings.

In relation to peer grouping, nine commenters stated that the Star Ratings provide unfair comparisons of hospitals that are fundamentally different.

- One commenter stated Star Ratings does not account for medical services and high acuity and vulnerable patient populations their medical centers disproportionately serve.
- One commenter explained that the ratings create disadvantages for high volume and large teaching hospitals that broad a variety of services.
- Two commenters noted that hospitals provide different services
 - One of the commenters also noted that hospitals report different measures.
- Two commenters noted that teaching hospitals, large hospitals, and hospitals serving low-income patients tend to receive lower star ratings, despite providing high-quality care.

- One commenter noted that teaching hospitals typically care for sicker and more vulnerable patients in diverse and complex environments.
- One commenter stated that Star Ratings are not apples to apples comparisons, and no incentive to report more measures. The commenter recommended consistent weights for measures to incentivize reporting measures.
- One commenter stated all hospitals are compared despite the number of measures they are scored on or the types of services they provide. The methodology does not expose differences in services provided, which makes it difficult for patients to make choices based on the care they need.
- Another commenter stated the methodology does not shed light on what services a hospital provides, and patients lack the information needed to make choices based on measures and services hospitals actually provide.
- One commenter stated the methodology biases large tertiary care centers and hospitals caring for more disadvantaged patients. They added that analyses performed indicate hospital size may impact ratings, as small hospitals tend to cluster in the middle and large hospitals are more spread to the ends of the Star Rating categories.

Twelve commenters provided additional comments on the overall methodology.

- Nine commenters supported condition-specific or clinical topic area Star Ratings.
- Another commenter recommended developing ratings for subsets of measures, to provide more meaningful information to consumers and hospitals. The commenter added that currently many measures are included in the rating, but only some ultimately impact the overall rating.
- One commenter requested re-examination of the underlying methodology to improve reliability, predictability, and accuracy. The commenter noted that the Star Ratings should be useful to consumers, stable and accurate, and show a clear line from performance on measures to Star Ratings. They added that CMS should take strategic steps to ensure confidence by all stakeholders in the Star Ratings.
- One commenter encourages CMS to revisit Star Ratings methodology to improve reliability, predictability, and accuracy.

Response: Thank you for your comments. CMS agrees that the Star Rating should convey reliable and useful information, through application of a rigorous statistical methodology. The current methodology was developed and vetted through consistent stakeholder engagement, including TEP and work group meetings as well as multiple public comment periods. The July 2019 confidential preview period highlighted the sensitivity of the methodology to changes in the underlying measures and CMS is prepared to further evaluate and refine the methodology as measures and stakeholder needs evolve. CMS recognizes that small hospitals have measure reporting patterns that differ from large acute care hospitals, and CMS will continue monitoring the interpretation and use of Star Ratings and will continue to investigate areas for improvement.

Latent Variable Modeling

Eighteen commenters provided general feedback on the LVM used within the Star Ratings methodology.

- One commenter suggested adapting the LVM to make it more predictable.
- One commenter noted that while they conceptually prefer the LVM, they expressed concern over CMS's ability to improve its sensitivity, validity, and reliability.
- One commenter noted the LVM approach produces results that are neither reliable nor reproducible. The commenter noted this leads to unstable ratings that change substantially.
- One commenter expressed concern over the reproducibility, bias modeling and reliability of the LVM.
- One commenter noted that the use of LVM introduces variability and inconsistency, making the ratings hard to interpret. The commenter further noted that LVM causes confusion, despite goal to provide clear information.
- Another commenter recommended removing disproportionate weighting on certain measures caused by LVM that cancel out performance of measures.
- Another commenter stated their facility is impacted by the lack of transparency and stability of LVM. The commenter added that the model makes it hard to predict performance, as shown by the loading changes in PSI-90 between reporting periods.
- Two commenters stated the LVM creates loading factors for measures that leads to measures being disproportionately weighted. The commenters added that the shift in loadings for the Hip/Knee complications and PSI-90 measures between December 2017 and July 2018 caused drastic shifts in Star Ratings despite national performance changing much, and the LVM remains overly sensitive to subtle changes in the underlying data.
- Another commenter had concerns about the application of LVM for measure group scores, adding that they feel the models have produced grossly fluctuating coefficients. The commenter said that a metric important enough to merit inclusion should positively influence performance. They said that the LVM assumes group variables correlate with each other as manifestations of a latent factors. They added they are concerned that this underlying assumption is incorrect.
- One commenter stated the LVM is inappropriate for this type of data, and analyses have demonstrated that the modeling approach causes large swings in measure loadings even when measures should be stable and large changes in Star Ratings due to statistical modeling rather than change in performance. The commenter noted the negative loadings produced in July 2018 which can penalize hospitals for good performance. The commenter recommended abandoning LVM and instead applying consistent weights for measures that can be evaluated annually in order for the Star Ratings to have value to the public.
- Two other commenters stated that ratings are dependent on few measures with narrow aspects of hospital care, while other broader measures hold little importance in the ratings. This gives providers inconsistent and unclear signal about where to focus quality improvement.
- Two other commenters stated that the LVM methodology should be revised or replaced.
- One commented that the LVM is rather complex and Star Ratings would benefit from less sophisticated methodologies.
- Two commenters stated that the LVM within the methodology is confusing and unreliable.
 - One commenter noted that other methodology updates will not be valuable until issues with LVM are addressed.

- One commenter offered to collaborate with CMS to determine effective solution to modify and improve the Star Ratings.
- One commenter noted that the current LVM method is complex and encourages CMS to use a straight forward weighting method.
- Two commenters encourage CMS to reconsider the LVM methodology and elaborates that a more rational approach for addressing measurement precision in Star Ratings is needed for accuracy, stability, and balance.
- One commenter noted the LVM introduces inconsistency between these CMS reporting programs, and emphasizes PSI-90.
- One commenter recommended a methodology approach presented in a publication titled “An Efficient Frontier Approach to Scoring and Ranking Hospital Performance”, to replace LVM.
- Two commenters noted the LVM results in imbalances measure loadings.
 - One commenter suggests that there is no reason for the PSI-90 measure or the Hospital-Wide Readmissions measure to have imbalanced loadings.
 - One commenter explained their concern about the imbalanced contributions of different measures and noted the LVM methodology to be instable. They further explained how a measure can *overfit* the model and notes that this has happened for the readmissions measure, PSI-90 and Hospital-Wide Readmissions.
- Eleven commenters attributed the period-to-period shifts in Star Ratings to LVM methodology.
 - Three commenters stated the current latent value methodology is inconsistent and too sensitive to subtle data changes, which can result in large changes in hospitals’ star ratings seemingly uncorrelated to changes in hospital performance and/or national performance.
 - One of the commenters noted this inconsistent reporting makes it difficult for providers to focus on areas for quality improvement.
 - One commenter noted LVM lacks transparency and disproportionally and inconsistently weights measures within groups. Since measure weighting cannot be predicted, this creates instability in the program.
 - One commenter noted significant shifts in Star Rating suggest the measures may not be weighted appropriately.
 - One commenter noted LVM is not the best approach to this data.
 - One commenter noted intertemporal smoothing would obscure instability of the LVM, but not fix it. A small change in the data can result in a large change in the measure weights, which can result in a hospital that improves in every dimension having a low star rating. This commenter proposed an alternative approach where the Star Ratings would better reflect the hospital data.
 - Four commenters noted that period-to-period shifts in Star Ratings are likely a function of shifting measure weights or “loadings” derived from the LVM.
 - One commenter specifically noted the change in loading values for PSI-90 and Hip/Knee Complications, which were weighted heavily in the Safety of Care measure group.

Response: The decision to implement Star Ratings with the current LVM methodology was based on a series of research, analytic work, policy considerations, and numerous stakeholder engagement

activities. However, as outlined within the public comment materials, CMS acknowledges that there are potential modifications and other methods altogether worth exploring to address limitations of the current methodology. We appreciate your comments and input included in this summary report on potential modifications to the LVM, such as incorporating measure precision, as well as alternatives to LVM, such as an explicit approach. Please note that with any methodology, including LVM and any alternatives, there are inherent advantages and disadvantages that contribute to CMS decision-making. We will consider your comments.

Stakeholder Engagement

Nineteen commenters provided input on stakeholder engagement efforts

- Seven commenters encouraged CMS to continue to engage stakeholders during reevaluation of the Star Ratings.
- One commenter proposed communicating any methodological changes (such as changes in measure weighting) to hospitals prior to reporting results.
- One commenter applauded the Patient & Advocate Work Group engagement and highlighted their priorities for meaningful and easily understood information, most current information, and avoiding confusing or misleading information. The commenter encouraged CMS to expand consumer engagement in the reevaluation of Star Ratings and *Hospital Compare*.
- One commenter recommended that CMS review how patients and providers currently use *Hospital Compare* and whether the proposed methodology updates meet those needs. Current research suggests that patients seek quality information for specific conditions or procedure.
- Four commenters recommended that CMS seek independent, impartial review of the methodology. Two of those commenters also requested consensus-based recommendations.
- Five commenters urged CMS to engage experts on LVM to ensure its calculation approach is executed correctly.

Response: We appreciate these comments. Engaging and being responsive to stakeholders are guiding principles of the Star Ratings. CMS has sought to elicit feedback from a range of stakeholders during the development and ongoing reevaluation of Star Ratings. CMS will continue to solicit stakeholder input as a collaborative effort to improve the Star Ratings for hospitals and consumers.

Measures Included in Star Ratings

Two commenters recommended decreasing the number of measures within Star Ratings to simplify the methodology and focus hospital improvement efforts.

- One commenter stated that it is difficult for facilities to have an effective strategy to address more than 15 measures.

Seven commenters noted differences among the measures included in Star Ratings.

- One commenter pointed out that the measures within Star Ratings vary in terms of whether and how they incorporate risk adjustment.
- Two commenters noted that the measures reflect different hospital settings (inpatient, outpatient, and emergency department) and varying impact on the final score.

- One commenter noted that the measures cover a variety of procedures and conditions.
- Three commenters stated that the measures within Star Ratings have different data collection periods
 - Two of the commenters noted that the different collection periods make it difficult for hospitals to pinpoint which measures are driving measure group performance.

Seven commenters provided general input on the measures included within Star Ratings.

- One commenter noted that the measures included in Star Ratings are mainly inpatient measures, which may not reflect the quality of care provided by hospitals primarily providing outpatient services.
- Three commenters stated that the current measures included in Star Ratings are disparate, lack clinical nuance, and oversimplifies complex factors that go into providing care.
 - One commenter added that it is important to ensure star ratings are not misleading due to flaws in the underlying measures.
- One commenter noted any rating system is only as good as the underlying measures. The commenter noted there are several issues with the reliability and validity of some of the individual measures included in Star Ratings.
- Another commenter suggested that available measures may not reflect services and treatments that matter to patients, that Star Rating might make inappropriate assumptions about patient priorities, and that CMS consider the types of measures that will provide meaningful results to patients and account for factors affecting hospital performance.
- One commenter suggested reconsidering the measures included in Star Ratings based on how the measures relate to overall quality provided by a hospital, per the healthcare consumer.
- Another commenter stated that the individual measures should be based on actual data rather than predicted data. The commenter stated that adjustments based on volume artificially adjust measure results, especially when measures are based on lower volume of patients. The commenter added that even if zero central line-associated bloodstream infections (CLASBI) events are reported, the score is adjusted towards the prediction that there may have been CLABSI to report if we more patients had been seen.

Seven commenters recommended only including valid and reliable measures in Star Ratings.

- One commenter further noted that changes in Star Rating should be driven by performance, not methodology, and measure loadings give a false impression of quality performance within the group.
- Other commenters specifically recommended including only NQF-endorsed measures.

Five commenters proposed including additional measures within Star Ratings.

- One commenter proposed including patient-reported outcome measures which capture mobility, mental status, and overall well-being.
- One commenter requested the addition of a “Patients who reported that their referral visit to a specialist was scheduled or confirmed prior to leaving the ED or hospital” measure to Patient Experience and a “Percentage of patients who are referred to a specialist or primary care provider for appropriate follow-up visit after an ED visit that did not lead to hospitalization”

measure to Effectiveness of Care, saying these processes can significantly increase patient compliance while improving readmission and complication rates and health outcomes.

- One commenter requested including MSPB as a measure of efficiency in Star Ratings.
- One commenter recommended including measures from other programs and registries, such as the National Cardiovascular Data Registry (NCDR) and Get With The Guidance (GWTG) registries in Star Ratings, and weighting them more heavily than billing data measures. The commenter recommended requiring hospitals to report standardized clinical registry data, instead of using data meant for billing as the gold standard. The commenter added that it is more expensive for hospitals to collect, but would shed light on which hospitals cares about improving quality.
- One commenter recommended incorporating quality measures beyond those reported on *Hospital Compare* into Star Rating, such as AHA structural and process measures.

One commenter provided general input on the measures included within Star Ratings.

- One commenter stated that data being publicly reported should be proven relevant and accurate for the type of facility being measures. They recommended not including a measure in Star Ratings if it does not demonstrate reliability.

Eighteen commenters provided input on specific measures, other than PSI-90 included within Star Ratings. See page 9 of Measure Grouping for specific comments on PSI-90.

- Two commenters recommended removing the HCAHPS survey from the methodology.
 - One commenter stated a Safety Net hospital could never attain the same patient satisfaction as other hospitals. The commenter noted that just because a hospital is noisy at night does not reflect poor quality of care or patient outcomes.
 - One commenter referred to studies demonstrating a positive association between HCAHPS scores and patient mortality rates and said that the HCAHPS survey scores have contributed to the opioid epidemic.
- One commenter noted that the Hospital-Wide Readmission measure is highly correlated with the Readmission measure group and contributes more to the measure group score than the eight other measures. The commenter stated that tertiary care hospitals accepting critically ill, high-acuity outlier patients are at risk for lower readmission measure performance, which impacts their Readmission measure group score and star rating. The commenter stated, based on internal analyses, that unavoidable readmissions for four patients resulted in their February 2019 star rating decrease.
- Three other commenters recommended removing readmission measures from Star Ratings.
 - One commenter stated they should be removed until they are adjusted for social risk factors. They stated that the lack of adjustment is a concern for their hospital which serves a disproportionately vulnerable population of patients facing challenges upon discharge.
 - One commenter stated the measures go beyond what a hospital itself can control and instead reflects factors like patient compliance, transportation availability, and access to community-based follow-up care.

- One commenter recommended CMS remove HRRP measures from Star Rating until the effect of the HRRP is better understood, citing some concerns about the benefits of HRRP.
- Two commenters provided input on the Sepsis-1 measure.
 - One commenter stated the Sepsis measure is controversial and should not be publicly reported until the algorithm is worked on to make the sepsis bundle data points more straightforward.
 - One commenter agreed with the incorporation into Star Ratings but recommended separating the measures rather than including the bundle. The commenter added the individual measures allow for organizations to more easily track performance and improvement.
- One commenter recommended CMS remove PSI-90, PSI-4, and the Hip/Knee complication measure from Star Rating due to concerns about the accuracy of the data used.

Response: Thank you for your comments. The Star Ratings is meant to summarize and be inclusive of the existing measures publicly reported on *Hospital Compare*. Since first reported in 2016, the Star Rating methodology has excluded several types of measures: measures with no more than 100 hospitals report performance publicly; structural measures (with no evidence of an association with improved outcomes); non-directional measures; measures not required for public reporting; and measures overlapping with another included measure. Additionally, beginning in February 2019, any measures with a statistically significant negative loading would be excluded. Any measure publicly reported on *Hospital Compare* not meeting any of these criteria is included in Star Rating, in accordance with the project's founding principle of measure inclusivity. The Star Rating methodology is additionally intended to accommodate changes and evolution in the underlying measures over time while maintaining methodologic rigor. Any measure-level updates are reflected accordingly within Star Ratings.

Individual measure methodologies and results are within the purview of individual measure stewards. Each measure has been vetted through standard processes to ensure valid measure concepts and reliable results, such as convening TEPs, soliciting public comment, pursuing NQF endorsement, and rulemaking.

We acknowledge that the Hospital-Wide Readmission and PSI-90 measures are assigned high measure loadings within the Readmission and Safety of Care measure groups, respectively. LVM assigns high measure loadings to these measures because the measures capture information on many patients and therefore carry a more substantial quality signal. However, as outlined within the public comment materials, CMS is exploring approaches to make the measure loadings more balanced within measure groups and consistent over time in response to stakeholder concerns.

Measure Group Weights

Nine commenters provided feedback on measure group weights.

- One commenter recommended decreasing the weight of the Safety of Care group due to the observed randomness, and increasing the weight of more stable domains.
- One commenter stated that it has been difficult to establish correlation between readmission rates and other traits of high quality care. The commenter noted several studies which show

conflicting results regarding correlation, and because of the ambiguity, recommended decreasing its contribution weight in Star Ratings (and increasing mortality weight).

- Three commenters recommended the mortality group be weighted more heavily.
 - One commenter acknowledged the weights have been vetted through stakeholder groups, but did not agree with Readmission and Mortality groups being assigned equal weights. The commenter suggested re-vetting the weights.
 - Another commenter recommending increasing the weight of mortality and reducing the weight of safety and readmission groups. The commenter noted the mortality measures have clinical importance, have been proven valid and reliable, and lack the possibility of surveillance bias that other measures such as PSIs are vulnerable to.
 - The commenter added that readmission measures are inherently less important than mortality measures because readmission is a better outcome than death, and many times planned.
 - Another commenter noted there are hospitals who receive high star ratings while performing worse than the national average in the mortality measure group.
 - Four commenters added that measure groups and group weights should be balanced and reflect importance to patients.

Response: Thank you for your comments. Most stakeholders agreed that outcome measures, such as mortality and readmission, should be weighted more than process and efficiency measures. We will continue to evaluate the weighting of measure groups in relation to other methodology updates and as the measures on *Hospital Compare* evolve.

Incorporating Socioeconomic Risk Adjustment

Twenty-four commenters recommended incorporating socioeconomic risk adjustment in Star Ratings.

- One commenter noted the methodology favors suburban hospital who serve a less vulnerable population, so socio-economic risk adjustment should be factored in. The commenter suggested peer grouping by large teaching, small teaching, small community, large community, critical access or specialty designation.
- One commenter noted that when socio-demographic status is not incorporated into the methodology, hospitals with a higher proportion of complex patients have lower hospital star ratings. The commenter encouraged monitoring this potential unintended consequence, and continue to look for ways to adjust for the risk for social risk. The commenter added that both academic medical centers and safety net hospitals who care for a more complex patient population are disadvantaged, and socio-demographic status risk adjustment allows for more fair comparisons.
- One commenter requested a fairer approach to account for social determinants and hospitals who serve a broader array of patients and services.
- One commenter stated that research has demonstrated clear differences in Star Ratings based on HRRP socio-economic status peer group assignments. Star Ratings, or at least the Readmission group, should be adjusted for socio-economic status.
- Another commenter noted studies showing the relevance and importance of sociodemographic factors, and recommended adjustment of readmission measures used in Star Ratings.

- One commenter noted the efforts by the NQF in the Social Determinants of Health Data Integration Project to account for differences in socioeconomic status among patients. The commenter stated Star Ratings and any quality reporting program should work with NQF in this effort.
- One commenter stated Star Ratings do not risk adjust for patient's socioeconomic and demographic circumstances, when we know these have a direct impact on a patient's health outcome.
- One commenter stated adjustment for social determinants of health are necessary for fair and meaningful comparisons of hospitals.
- One commenter stated that evidence suggests the rating system disproportionately impact safety-net hospitals and large hospitals that report additional measures. The commenter added that the methodology does not account for social determinants of health, although CMS has publicly acknowledged the impact social factors have on quality measures. The commenter stated and presented analyses to show a relationship between the number of stars awarded and socio-demographic factors (hospital zip code and patient case-mix levels), and between the number of measures reported and domains used in the models.
 - The commenter added that the model is sensitive to measures with questionable validity and contains redundant constructs in the readmission group.
- One commenter noted variation in social risks among different communities served should be accounted for.
- One commenter recommended risk adjustment methodologies that account for patient social determinants of health, insecurities, race, ethnicity, and education, as well as the complexity of services provided by a hospital which all impact readmission, patient satisfaction and PSIs.
- Four commenters requested accounting for social risks that are outside the hospitals' control, as other CMS programs do. The commenters noted that two-thirds of the star rating summary is based on mortality, readmission, and patient experience which have all been shown to be influenced by social risk factors.
 - One of the commenters added that their hospitals provide care beyond medical treatment for disadvantaged patients, such as ensuring discharged patients have nutritious foods. The commenter added that research shows factors outside of the hospital's control influence readmissions, and called for risk adjustment (including sociodemographic status, language, and post-discharge support structure) to ensure results are accurate and reflect varying patient characteristics across hospitals. The commenter identified the Medicare Advantage star rating as example.
- One commenter recommended risk adjusting for socioeconomic status and poverty, as well as aligning volume/size facilities providing similar services. The commenter noted recent research has shown hospitals located in lower socioeconomic areas receive 1 to 3 stars, and HCAHPS scoring is lower due to low socioeconomic status.
- One commenter recommended a combination of peer grouping and risk adjustment for social determinants of health to address differences in academic and safety net hospitals, and transfer patients.
- Another commenter supported peer grouping in conjunction with a risk adjustment methodology so that hospitals with similar characteristics and risk profiles were compared to each other.

- One commenter stated SES analyses indicated as lower SES patients increased, star rating decreased and vice versa.
- One commenter recommended peer grouping star ratings using socio-economic factors (e.g., income, age, education, employment, uninsured and housing) among similar hospitals.
- One commenter requested social risk adjustment in all measure groups, as academic medical centers have a larger amount of complex medical treatments and tend to have larger denominators on measures that are more complex (such as organ transplants and surgeries) compared to small community hospitals which tends to have a larger denominator in measures such as pneumonia. The commenter added that the current methodology rewards community hospitals for not performing certain types of care.

Response: Many commenters recommended direct socio-economic status risk adjustment, which CMS has acknowledged as an important topic and consideration in the context of quality measurement, when appropriate. Each measure is being assessed for socio-economic risk adjustment individually during NQF endorsement and maintenance. Currently, the purpose of the Star Rating is to summarize existing measures, as they are reported on *Hospital Compare*. Socio-economic risk adjustment may not be appropriate or warranted for all measure types, for example the rate of healthcare-associated infections, and therefore make it potentially inappropriate to adjust summary scores or star ratings. CMS will continue to evaluate individual measures and monitor their impacts on Star Ratings.

Alignment with Other Quality Metrics and Programs

Nine commenters suggested better alignment with other CMS programs or quality rating efforts.

- One commenter specifically recommended including the same measures within the HRRP and the readmission measure group. The commenter noted that the pneumonia, acute myocardial infarction, and heart failure excess days in acute care (EDAC) measures as well as the stroke readmission and hospital-wide readmission measures are included in the Star Ratings readmission measure group but not within HRRP. The commenter recommended that only measures within CMS programs, such as HRRP, be included in Star Ratings to best reflect hospital improvement efforts made for payment programs.
- Two commenters noted that there are discordant results between the Overall Star Rating and HCAHPS star rating, Leapfrog, HealthGrades, US News and World Report.
 - One commenter noted stakeholders and patients do not have access to expanded metrics to understand the difference in ratings. The commenter recommended that CMS provide a scorecard, similar to other rating systems, that are easier for the public to comprehend.
- An additional commenter acknowledged that the measures included in the ratings and other CMS programs are similar, however there is a lack of concordance in methodology. The commenter suggested tying incentives to the star rating measure groups, which would align with MedPAC's recommendation to consolidate programs.
- One commenter stated that CMS sets the nation's standards for health care performance through their pay-for-performance programs, which include many measures that contribute significantly to the star rating, however the results between the programs and star ratings are inconsistent. The commenter attributed the differences to differences in methodologies (such as HRRP using quintiles of proportion dual-eligible patients and star ratings not using this

adjustment), and underlying measure importance (for example HAls not having significant importance in star ratings but affecting payment in other programs), and requested better alignment in methods used. The commenter added that this hurts hospital reputation and creates confusion for consumers.

- Another commenter recommended the Star Rating methodology produce results more similar to performance in other CMS programs including HVBP, HRRP, and HACRP. The commenter noted that more aligned results will improve general acceptance of the rating system. The commenter supported exploring explicit methodologies, such as those used in other CMS programs, to produce more easily understood results.
- One commenter also recommended defining measure weights similar to how they are defined in other CMS programs.

Five commenters stated that the Star Rating is inconsistent with research trends, national and state quality metrics, other pay for performance programs, or other private quality ratings.

- One commenter stated that even though the same measures are used across CMS programs and Star Ratings, the results are inconsistent, especially for safety measures.
- One commenter noted that California had already created a state-level star rating system and multiple star rating efforts with different results confuse patients.

One commenter provided other comments on the alignment with Star Ratings and other quality metrics programs.

- One commenter recommended a methodology similar to Truven/Watson Top 100 Hospitals.

Response: Thank you for your comments. While alignment with other CMS programs is a guiding principle of Star Ratings, each program includes different measures and has different approaches to aggregating quality information as well as different purposes and incentives. CMS will continue to explore reasonable ways to align the Star Ratings with other payment programs and star rating efforts.

[Display of Star Ratings](#)

Nine commenters provided input on the resources and display of Star Ratings.

- Five commenters commented on the current display of Star Ratings on *Hospital Compare*.
 - One commenter provided input that the public should be able to understand the rating system at a fourth-grade level. The commenter suggested showcasing expanded metrics for metrics of interest, and not showcasing negative metrics. The commenter noted this view would allow consumers to find the best hospital for services of interest.
 - One commenter recommended changing the Star Rating display on *Hospital Compare*. The commenter stated that the rating being displayed on every page gives consumers the incorrect impression that the displayed star rating on the page refers to the specific metrics displayed.
 - One commenter recommend CMS add a notice to consumers that Star Ratings are only one aspect to consider when choosing a provider. The commenter also recommended including the cut-points that distinguish each star category.
 - One commenter shared that their hospital has not had enough data to receive measure scores, and therefore a star rating, and expressed concern that the public may interpret

“N/A” in place of a star rating as failing to participate or perform. The commenter elaborated that they worry this would negatively impact applications for insurance carriers, future referrals, patient volumes, and reimbursement. The commenter recommended that either CMS consider approaches that would account for hospitals without measure information because of size or volume or add language on *Hospital Compare* as to why a hospital would receive a N/A rating.

- One commenter recommended the preview period be longer than 30 days.
- Six commenters questioned the current consumer use of the Star Ratings and *Hospital Compare*.
 - One commenter requested CMS examine whether the public is accessing Star Ratings online. The commenter noted another option would be to show specific measures and aggregate data into ratings based on specific patient populations.
 - Another commenter questioned whether consumers of various demographics use *Hospital Compare* and Star Ratings, or if it is mainly hospitals using the site and ranking.
 - One commenter requested CMS publish objective measures to assess whether the Star Rating is working as intended for patients and health systems. The commenter suggested a patient-level measure include an assessment of how often Medicare beneficiaries are using the Star Ratings in health care decisions and a health system-measure how well health systems believe the Star Ratings reflect the true clinical quality of care provided at their hospitals.
 - One commenter pointed out that, despite summarizing many different measures, most hospitals receive a three-star rating, making most hospitals appear average. Therefore, the ratings have limited utility for patient decision-making and incentives for performance improvement. They recommended allowing consumers to view the summary score that informs the star rating to provide more meaningful differentiation among hospitals.
 - One commenter also requested CMS examine if the public is accessing Star Ratings online. The commenter noted another option would be to show specific measures and aggregate data into ratings based on specific patient populations.
 - One commenter urged CMS to create more user-friendly website.
- Eight commenters provided suggestions about supplemental resources for stakeholders.
 - One commenter suggested CMS publish full details of its model for hospitals to anticipate new ratings.
 - One commenter suggested that CMS provide feedback to the hospital or organization about why their scores changed.
 - Another commenter recommended a more comprehensive explanation of how the Star Ratings are calculated.
 - Two commenters requested a guide or roadmap resource to assist facilities in improving their performance on the measures that are included within Star Ratings.
 - Another commenter recommended creating a statistical toolkit that provides test data to allow for developers to recreate models in other statistical packages, provides crosswalks from the Virtual Research Data Center (VRDC) to SAS code inputs, and

provides a pathway for proposed revisions to the methods via an open source method. This type of toolkit would accelerate advances from the community.

- One commenter also suggested investigating additional strategies to make it easier for health systems to understand recent performance of peer institutions on clinical outcomes contained in the Star Ratings.
- One commenter urged CMS to host education sessions to help hospitals understand how to improve measure scores and set achievable goals, and develop a transparent, reproducible methodology. The commenter suggested publishing deciles or quartiles and benchmarks rather than results that cannot be reproduced. For example, The Health Services Cost Review Commission provides the calculation for quality-based reimbursement equations to hospitals prior to reporting.

Response: Thank you for your comments. At the inception of Star Ratings, the display of Star Ratings and accompanying text was determined based on website capabilities as well as through stakeholder input. CMS conducts ongoing consumer testing of *Hospital Compare* as a whole. The Star Ratings display can be reevaluated as website capabilities and consumer needs evolve. Although beyond the scope of CORE's work on the Star Rating methodology, CMS continues to evaluate the *Hospital Compare* website in general, and resources included, as more capabilities become available. CMS will consider your suggestions as the webpage continues to be evaluated for clarity and usefulness.

CMS is committed to educating and supporting all stakeholders, including consumers and hospitals, through comprehensive resource materials, question and answering support through an email inbox, and national provider calls. In addition, each hospital receives a confidential, hospital-specific report, which outlines their star rating, measure group performance, and individual measure performance. This resource is intended to maximize transparency and allow hospitals to view the measure-level information that contributes to their star ratings. The model specifications, in the form of a methodology report, are posted publicly on the QualityNet website concurrent with the confidential, hospital-specific report so that hospitals can review and interpret their Star Rating results. The Star Ratings data, SAS pack, and supplemental documentation guide for a given reporting period is publicly posted at the time of public reporting on *Hospital Compare*.

Comments Beyond the Scope of Star Ratings Project

We received two comments from consumers about personal healthcare experiences. While their comments are beyond the scope of the Star Rating methodology, the comments express a need for continued public reporting of hospital quality information that is useful for patients and consumers as they navigate the healthcare system.

Response: Thank you for your comments. We appreciate the consumer comments and acknowledge the need for useful public quality information.

Two hospitals commented on their hospital-specific Star Rating.

- One commenter requested a review of their hospitals Star Rating, as they do not believe it reflects the care provided to patients.
- One commenter noted that in 2017 and 2018, their hospitals went from receiving a 5-star to not receiving a star rating due to too few measures reported. The commenter noted this was

because their HAI measures were listed as N/A because the score was less than one. They noted this penalized their hospital for good performance.

Two commenters provided feedback on the Hospital Quality Reporting (HQR) platform and *Hospital Compare* in general.

Thirty-six commenters provided feedback on the individual measure methodologies or applications to other CMS programs.

Twenty commenters offered feedback on measures in general, measures across multiple groups, or other measurement programs.

Response: Thank you for your comments. Your input on individual measures or other CMS programs has been provided to CMS, however your comments are beyond the scope of the Overall Hospital Quality Star Ratings, and therefore not elaborated upon within this summary document. The Star Ratings project is meant to summarize and be inclusive of the existing measures publicly available on *Hospital Compare*. It is beyond the project scope to evaluate individual measures or to adjust individual measure scores. Individual measure methodologies are within the purview of individual measure stewards. Please note that the individual measures have been themselves extensively developed and evaluated. In accordance with the Star Rating project's founding principle of measure inclusivity, it is assumed that every publicly reported measure has met certain criteria of validity and reliability and is therefore eligible for inclusion in the overall summary rating. For commenters concerned about individual Star Ratings, inquiries may be sent to Lantana Consulting Group at the cmsstarratings@lantanagroup.com inbox.

Overall Analysis of the Comments and Recommendations

The Star Ratings are designed to be a summary of the existing publicly reported measures on *Hospital Compare*. The topics outlined within the public comment materials were identified in response to previous and current stakeholder concerns about consistency and comparability. CMS will consider improvements to the Overall Star Ratings methodology based on comments received from past stakeholder engagement efforts and the most recent public comment period. Other questions or comments can be sent to the following email address at any time: cmsstarratings@lantanagroup.com. CMS is committed to evolving the Star Ratings methodology through reevaluation work and stakeholder engagement.