AHPA
ADVENTIST HEALTH
POLICY ASSOCIATION

March 29, 2019

VIA ELECTRONIC MAIL
cmsstarratings@yale.edu

Kate Goodrich, M.D.
Chief Medical Officer
Center for Medicare and Medicaid Services
7500 Security Boulevard
Baltimore, MD 21244

Re: Overall Hospital Star Rating on Hospital Compare Public Input Request

Dear Dr. Goodrich:

On behalf of the Adventist Health Policy Association (AHPA) we appreciate the opportunity to comment on the proposed methodology enhancements for the Overall Hospital Star Rating System. Our organization of five Seventh-day Adventist affiliated health systems includes 84 hospitals and more than 300 other health facilities in 17 states and the District of Columbia. AHPA represents a major segment of the U.S. hospital sector. Our member hospitals operate in a variety of settings, ranging from rural Appalachia to urban areas of California.

Below, please find AHPA's comments and recommendations in response to the Public Input Request. Specifically, we comment on the following issue areas:

- Measure Grouping
- Regrouping of Measures
- Incorporating Measure Precision
- Period-to-Period Star Ratings Shifts
- Peer Grouping
- Closed-Form Solution
- Explicit Approach to Calculating Overall Hospital Quality Star Ratings
- Alternatives to Clustering
- Incorporation of Improvement
- User-Customized Star-Ratings

General Comments

In 2015, the Centers for Medicare and Medicaid Services (CMS) introduced star ratings on Hospital Compare, the Agency's public information website, to make it easier for consumers to choose a hospital and understand the quality of care that hospitals deliver. However, the methodology currently used to calculate the star ratings has led to inconsistencies and made it difficult for hospitals to predict their score. To reevaluate the Overall Hospital Quality Star Rating on the Hospital Compare website, CMS contracted with the Center for Outcomes Research and Evaluation (CORE). CORE seeks public input on their proposed methodology enhancements.

We commend CMS' resolve to improve the usability, accessibility and interpretability of Hospital Compare for patients and consumers. While we support CMS refining the Overall Hospital Quality Star Rating, we have some general concerns and suggestions surrounding CMS' proposals that were included in the public input request. Our comments can be found below.

Methodology Enhancements by Category

Measure Grouping

CMS is reconsidering how it groups measures and defines measure groups. CMS is seeking public feedback on its three-step approach to regrouping, including separating the Safety of Care group into two separate groups. Additional areas for discussion include removing the denominator weighting. AHPA's response to the questions on section 4.1 on measure grouping is as follows:

CMS would like to use a three-step approach (clinical coherence, confirmatory factor analysis, and ongoing monitoring) to define measure groups. Is this approach reasonable?

AHPA believes that the proposed three-step approach to define measure groups is reasonable. However, we recommend that CMS use quantitative instead of qualitative criteria to determine and evaluate the coherence, strength, balance and consistency of measure groups. We also encourage CMS to provide further clarification on the process that would be used to determine clinical coherence when defining measure groups.

CMS proposes, as part of the proposed three-step approach, to conduct a confirmatory factor analysis that would help examine if a group of measures can be explained by a single common underlying factor. CMS decided to use the criterion of "ratio of first to second eigenvalue in weighted factor analysis greater than three" as a guide to identify the dominance of one factor. The Agency indicates that a larger first eigenvalue is desired because it indicates that a single underlying factor is strongly associated with all measures in the group.

AHPA believes that the qualitative assessment of the shape of the eigenvalue screen plot can lead to the potential inaccurate inclusion of a group. Consider Figure 3 and Figure 4 of the Public Input Request. Figure 3 passes the ratio test of greater than three. CMS notes that Figure 3 demonstrates a well-constructed mortality group because it displays a prominent turning point at the second eigenvalue, whereas Figure 4 for the safety group does not. However, the prominence of the turning point is dependent on the third eigenvalue and may be on subsequent ones. If the third eigenvalue is closer to the second eigenvalue, the kink is more pronounced even if the ratio is below three. Similarly, if it is much below the second eigenvalue, the kink is flatter even if the ratio is much higher than three. In this situation, it is not clear what a qualitative assessment would find.

CMS also proposes to conduct ongoing active monitoring to ensure that measure loadings are balanced within each group and relatively consistent over time. While we commend such effort, there is no clear understanding of how CMS would determine the balance and consistency of new measures, especially for measures for which there is not enough historical data. **Based on this issue, AHPA recommends that CMS provide further guidance on how the Agency would determine whether a group is balanced and consistent.**

Should CMS use the balance and consistency of loadings as a factor in evaluating grouping?

The balance and consistency of loadings is crucial to evaluate measure grouping. However, we recommend that statistical criteria be used to determine thresholds.

Regrouping of Measures

CMS is considering regrouping measures, particularly within the Safety of Care group, and seeks input on how to reform the grouping of measures. The Agency notes that the measures in Safety of Care are

potentially contributing to the inconsistency and lack of predictability in the current methodology. AHPA's response to the question in section 4.2 on the regrouping measures is as follows:

Is the current grouping or one of the potential alternative groupings of the Safety of Care measures most suitable for the Overall Hospital Quality Star Rating based on previously mentioned criteria?

AHPA believes that neither the current group nor the alternatives provided are suitable. In Option 2, the surgical safety group ratio of eigenvalues does not meet the conventional threshold of three and is very close to the current eigenvalue. Additionally, the group contains both Healthcare-Associated Infection (HAI) and Patient Safety Indicator (PSI) measures, which use different types of denominators that would cause skewed denominator distribution. This problem is currently present in the Safety of Care group and may also be a problem for readmission grouping, with the introduction of the Excess Days in Acute Care (EDAC) measures. **AHPA recommends using an alternative approach to categorize measures into groups by performing factor analysis on different permutations of measures without a priori groupings.** This could be done on all measures but especially on safety measures. An approach without a priori assumption might suggest more efficient and statistically sound grouping with all or some of the measures.

For the Safety of Care group, we also recommend removing the PSI-90 composite measure. Research has found many deficiencies with PSI-90, which include the following: susceptible to surveillance bias; may not be preventable through evidence-based practices; lacks appropriate and necessary exclusions; and are based on administrative claims data so cannot capture the full scope of patient-level risk factors.[1] The current PSI-90 measure is also hospital-specific and does not capture harm throughout the entire continuum of care. We believe that CMS should invest in methodologies that capture harm in a more accurate and comprehensive manner. This should involve capturing data that reflects the patient experience across the full health care continuum, including the hospital, ambulatory and post-acute care settings.

[1] Rajaram, Ravi et al. Concerns About Using the Patient Safety Indicator-90 Composite in Pay-for-Performance Programs. JAMA. Vol 313, No. 9. March 3, 2015. Retrieved from: http://jama.jamanetwork.com/article.aspx?articleid=2109967

Incorporating Measure Precision

Currently, CMS accounts for differences in measure score precision by using denominator weighting. This method places a higher weight on measures with a larger denominator, given that larger denominators are correlated with precision. However, a recent assessment found that this method contributes to the imbalance of measures and worse model fit. To address this issue, CMS is proposing different weighting approaches: denominator weighting, confidence interval weighting, equal weighting, and log denominator for non-volume denominators. AHPA's response to the questions in section 4.3 on measure precision is as follows:

Do you have any concerns about changing the methodology to use a combination of denominator weighting and log denominator weighting, based on the type of measure?

AHPA believes that while none of the proposed weighting approaches completely address CMS' concern, log transformation of the denominator would be the most mathematically appropriate approach. Log transformation helps with normalization of the data anytime there are outliers or skewness. Other transformation methods should also be explored, depending on the distribution of a measure. Square-root and inverses or any Box-Cox transformation would likely remove skewness in the distribution. Additionally, other model performance statistics should be calculated for proper evaluation.

AHPA also recommends using the confidence interval weighting as an alternative that could be applied to all the risk-adjusted measures (Readmission, Mortality, Safety). Box-Cox transformations could be used for the non-risk adjusted measures to remove any non-gaussian attributes.

Do you have any concerns about applying a change to the weighting approach across all measure groups (where data are available) vs. applying the change only to measure groups that meet specific criteria?

AHPA believes that the weighting approach should be applied to measure groups that are not balanced or consistent.

Period-to-Period Star Rating Shifts

The refresh of star ratings from December 2017 to June 2018 caused many hospitals to see large changes in their ratings despite the methodology remaining the same. These large shifts in star ratings lead to

confusion among providers and consumers. To remedy this issue, CMS is considering adopting either a rolling average or an update once a year of the Overall Hospital Star Rating. AHPA's response to the questions in section 4.4 on Period-to-Period Star Rating Shifts is as follows:

What are possible benefits and drawbacks to increasing stability by incorporating data from an older period?

Although the benefits of using weighted averages can lead to less variation in performance and less variation to the star ratings, there is a drawback when hospitals achieve significant performance improvement in more current periods. The weighted average reduces the impact of that performance improvement on the star rating. Another drawback is that many of the measures are already based on old performance (three years back). Incorporating previous performance periods would expand the lookback period for hospitals' performance even more and would not be an accurate reflection of their current performance. This would make it more difficult for hospitals to make any substantive strides and achieve a higher star rating.

Should the Overall Hospital Quality Star Rating methodology be modified to incorporate data from previous periods through a time averaged approach?

AHPA believes that previous periods should *not* be considered in the calculation of star ratings. Star ratings should be updated annually after all measures refresh. This approach would provide more accurate and meaningful information to consumers.

Are there other approaches to this CMS should consider?

AHPA recommends including weighted averages for measures that contain no overlapping data and current data for measures with overlapping data. Ideally, all measures should be calculated based on current year data, rather than three-year lookback periods. Currently, measures are calculated over different periods, sometimes even within a group. To mitigate any adverse impact from this difference, consistency could be achieved by using the weighted average over the same period for measures with no overlapping data. CMS could also consider using a two-year lookback period for all measures. This approach provides more current data and can demonstrate a hospital's improvement sooner than the current three-year period used among several measures.

Should CMS only refresh the ratings annually instead of the current biannual refresh?

AHPA recommends that CMS refresh the ratings annually. Biannual ratings may cause confusion among consumers, particularly if they choose a facility based on its star rating to then realize that the rating changed within a couple of months.

Peer Grouping

CMS is considering calculating the Overall Hospital Quality Star Ratings by grouping hospitals into peer groups. For example, safety-net hospitals could be grouped together to generate a star rating. This would result in similar hospitals being compared to each other instead of every hospital being grouped together. AHPA's response to the questions in section 4.5 on peer grouping is as follows:

Would it be valuable to calculate Overall Hospital Quality Star Ratings among peer groups? How should the information be displayed? If CMS decides to move forward with this feature, which stakeholders do you believe would use the information and how would they use it?

AHPA supports using peer groups to calculate the Overall Hospital Quality Star Ratings as well as risk-adjusting for patient populations and complexity of cases. We believe that this approach, currently used in the HRRP, would more fairly compare performance among hospitals and lead to more accurate star ratings. Under the current methodology, hospitals with more data and reported domains tend to perform worse than hospitals with fewer measures and domains reported. Hospitals that report fewer measures (smaller, non-teaching, specialty hospitals) tend to have more stars. [2] These hospitals tend to be significantly different than their counterparts and a star rating system should account for these differences. Therefore, we recommend that CMS stratify the ratings based on both the characteristics and the types of hospitals. CMS should provide risk-adjustment for patient population and complexity of cases and use these components to cluster hospitals into different peer groups. AHPA urges CMS to risk-adjust measures for dual-eligible status. This risk-adjustment is currently being done in the Hospital Readmission Reduction Program (HRRP) and we believe it should be replicated across all Medicare quality programs. Alternatively, a hospital's patient population and the complexity of cases need to be factored into the structural equation models to calculate the proper loading scores for each hospital.

[2] Bilimoria KY, Barnard C. The New CMS Hospital Quality Star Ratings The stars Are Not Aligned. JAMA. 2016 Nov 1;316(17):1761-2.

Among the feasible variables that could be used for peer grouping (specialty, number of measures reported, teaching status, number of beds, critical access hospital, proportion of dual eligible patients), which would be most useful?

After adjusting for peer group characteristics, AHPA recommends using a single rating for all hospitals (an overall star rating) and a secondary rating based on peer grouping. We believe that this rating methodology would allow consumers to compare hospitals more accurately. For example, a patient interested in oncology services would be able to compare the quality of a cancer hospital with other cancer hospitals. **While AHPA does *not* recommend that the number of beds be used to establish hospital peer grouping, below are the variables that we *do* believe should be included:**

- Dual-eligible status,
- Type of hospital: teaching/non-teaching/boutique or specialty hospitals,
- Number of measures reported,
- Critical access or non-critical access hospital.

Computational Update: Closed-Form Solution

CMS proposes to replace the current computational "quadrature" approach, which solves the mathematical equations of the Latent Variable Model (LVM) and calculates the scores of measure groups with the "closed-form solution." CMS would adopt instead a "closed-form solution," which is expected to more quickly solve the equations and eliminate the time-consuming computations used in the quadrature approach. AHPA's response to the question in section 4.6 regarding the Computational Update is as follows:

Should CMS use a "closed-form solution" or make technical changes like this potential solution and consider opportunities for such changes in the future?

AHPA recommends that CMS find alternatives to the quadrature approach. We find the current approach to be cumbersome to implement.

Explicit Approach to Calculating Overall Hospital Quality Star Ratings

CMS seeks input on alternative approaches to the LVM that assign explicit weights to each measure in each group, independently of the performance distribution or relationships between measures. CMS

would like feedback on using an explicit approach, which could be implemented by assigning weights to each measure in each group and then calculating each hospital's group score as a weighted arithmetic average of its measure scores. AHPA's response to the questions in section 5.2 on the explicit approach is as follows:

What are the advantages and disadvantages of a more explicit approach to calculating Overall Hospital Quality Star Ratings?

The advantage of a more explicit approach is its simplicity and ease of understanding. LVM is not an intuitive model and the loadings may change over time depending on the data available at the time. Thus, hospitals cannot predict how they will perform on any measure or take any policy action to focus on specific set of measures. For these reasons, AHPA believes that the explicit approach would offer more transparency and predictability.

Is the explicit approach a worthwhile change in approach and direction to consider further?

Yes, AHPA believes it is the right course of action for CMS to take.

How could such an approach be best operationalized or sustained?

The determination of weights needs to consider measures that are more critical and thus require higher weights. While weights should be tailored to the measure set, they should also remain balanced across a measure group to avoid placing too much emphasis on any one measure. No one metric should be weighted much more or much less than the others.

Alternatives to Clustering

CMS seeks feedback on alternatives to the current k-means clustering that assigns each hospital to a discrete star rating category from the continuous distribution of summary scores. CMS is considering alternatives for grouping hospitals into star rating categories. AHPA's response to the question in section 5.3 on clustering alternatives is as follows:

Should CMS consider potential alternatives to k-means clustering in more detail?

AHPA believes that the current k-means clustering is statistically sound. However, the current method limits hospitals' ability to predict cut points in future periods and these time differences or lagged effects have an influence on the star ratings assigned. One option for CMS' consideration is to conduct cross validating with Support Vector Machine (SVM) if data has already been labeled by the rating system to improve the misspecification of borderline scores with k-means. Additionally, CMS could adjust ratings with partial stars after conducting the clustering analysis for hospitals near the cut-off points.

What other considerations should guide future CMS work regarding clustering?

AHPA believes that the purpose of clustering is to give a sense of the underlying probability distribution of the score. For this reason, methods that are sensitive to distribution shifts should be considered, such as the Mean Shift Clustering method.

Incorporation of Improvement

CMS is considering including a hospital's improvement in comparison to its own prior performance into the Overall Hospital Quality Star Rating methodology. AHPA's response to the questions in section 5.4 on the incorporation of improvement is as follows:

Should CMS consider incorporating improvement in future iterations of the Overall Hospital Quality Star Rating?

Yes, AHPA recommends that hospital improvements be considered in a hospital's star rating. AHPA suggests that this be done through a separate metric within the five-star rating, similar to the method in the Value-Based Purchasing (VBP) program. Both performance against peers and performance against self are both factors currently used in the Value-Based Purchasing (VBP) program. **However, AHPA urges CMS to move with caution when selecting a method to integrate previous data into a hospital's summary score for the purpose of demonstrating improvement.** Any method adopted should avoid penalizing consistently high performing hospitals for having less room for improvement than a low performing hospital.

What are conceptual benefits and risks of incorporating absolute score improvement into the Overall Hospital Quality Star Rating?

The benefit of incorporating absolute score improvement is increased information and transparency for patients. The risk is that higher-ranked hospitals may be perceived negatively by patients if they underperform from one year to the next.

User-Customized Star Ratings

CMS is considering creating a user-customized star rating tool. In this tool, the weights of the measures would be dependent on the consumer's preferences.

AHPA does *not* support the introduction of user-customization to the Overall Hospital Star Rating. Given the variability in patient population and preferences, it would be increasingly difficult for hospitals to develop targeted policies around any set of quality measures and achieve a higher star rating. If the explicit approach attempts to address provider concerns over predictability and interpretability of the weights, user-customization reverses it.

Furthermore, using a user-customized star rating tool might create a perverse incentive. If over time hospitals find a pattern in how their patients are using the customized ranking, they might focus on those measures more. For example, if Timeliness and Patient Experience are more heavily weighted by most of the hospital's population, those measures may pull focus away from other clinically important measures such as Mortality or Safety. **To encourage customization, an alternative approach would be to provide separate ratings for each of the measure groups in addition to maintaining the Overall Hospital Star Rating.** This would allow CMS to address different priorities, preferences or values among patients.

AHPA also believes that CMS should consider using an alternative approach to encourage the development of different hospital rating platforms by providing raw data to the private sector. Giving technology companies access to hospital quality rating data and offering incentives could lead to the development of software applications that meets the needs of the public.

Conclusion

AHPA welcomes the opportunity to further discuss any of the recommendations provided above. If you have any questions or would like additional information, please contact Carlyle Walton, President of AHPA, at Carlyle.Walton@AdventHealth.com or Julie Zaiback-Aldinger, Director of Public Policy and Community Benefit, at Julie.Zaiback@AdventHealth.com.

Sincerely,

Carlyle Walton, FACHE
President
Adventist Health Policy Association