American Hospital Association™
Advancing Health in America

Washington, D.C. Office
800 10th Street, N.W.
Two CityCenter, Suite 400
Washington, DC 20001-4956
(202) 638-1100

March 27, 2019

Kate Goodrich, M.D.
Chief Medical Officer
Centers for Medicare & Medicaid Services
7500 Security Blvd
Baltimore, MD 21244

RE: Overall Hospital Quality Star Ratings on Hospital Compare Public Input Request, February 2019

Dear Dr. Goodrich:

On behalf of our nearly 5,000 member hospitals, health systems and other health care organizations, and our clinician partners – including more than 270,000 affiliated physicians, 2 million nurses and other caregivers – and the 43,000 health care leaders who belong to our professional membership groups, the American Hospital Association (AHA) appreciates the opportunity to provide input on potential future changes to the Centers for Medicare & Medicaid Services (CMS) hospital overall star ratings system.

As longstanding supporters of transparency, America's hospitals and health systems believe that patients, families and communities should have valid, clear and meaningful quality information to help them make important health care decisions. That is why the AHA has strongly urged CMS to address the substantial flaws in the star ratings methodology since the ratings inception in 2016. We continue to be concerned that one of CMS's laudable goals with star ratings – to give a meaningful, simplified view of hospital quality to consumers – is being compromised by a methodology that can lead to inaccurate, misleading comparisons of quality performance.

The AHA appreciates CMS's ongoing efforts to solicit stakeholder feedback on how to improve the ratings approach. The roughly one dozen potential changes to the star ratings methodology outlined in the request for comment attempt to address several important issues with star ratings and merit serious consideration. **However, the AHA believes that only three of the proposals should be pursued further at this time – empirical criteria for measure groups, peer grouping star ratings among similar hospitals, and using an "explicit" scoring approach.** The remaining proposals either fail to address important shortcomings with star ratings, or simply do not have enough information for us to judge their impact.

The AHA also urges CMS to consider other steps to improve star ratings that are not addressed in the draft report. We believe it is important that these steps be taken prior to considering implementation of any other changes to the star ratings. Specifically, CMS should:

- Engage a small group of experts on latent variable models (LVM) to ensure its calculation approach is executed correctly.
- Examine how to mitigate the impact of outliers in calculating readmissions measures in the ratings.
- Develop an alternative approach to star ratings in which, instead of an overall rating, hospitals receive ratings on specific clinical conditions or topic areas.

Lastly, we continue to urge CMS to remove the existing star ratings from *Hospital Compare* while its important work of improving the methodology continues. We appreciate the desire for the ratings to reflect the most current quality data. Yet CMS's public comment underscores the many problems with the current methodology. Unless and until the ratings methodology is improved, it will be difficult for hospitals and the public to have confidence that star ratings portray hospital performance accurately.

Our comments below describe the elements that any approach to hospital star ratings must have in order to be a credible rating system. We then provide more detailed comments on the extent to which CMS's proposed changes address these elements, as well as comment on several other issues.

THE "MUST HAVE" ELEMENTS OF A STAR RATINGS SYSTEM

Since CMS began work on overall star ratings in 2015, the AHA has repeatedly shared with the agency our ideas and concerns about the star ratings approach. In general, our concerns have asked CMS to address what we believe are six "must have" elements for the design of any star ratings system. These elements are described in greater detail below.

Usefulness to Consumers. The ratings should provide information that is relevant to the wide range of reasons patients seek hospital care, and give consumers the ability to drill down on the particular aspects of care most relevant to them. As currently designed, we are concerned that star ratings do not reflect the aspects of care most relevant to a particular patient's needs. For example, a family may be interested in selecting the best hospital for cancer care, but there are no such measures included in the current star ratings.

Accuracy. The ratings should be based on rigorous quality measures, and employ appropriate and correctly-executed statistical approaches to combining performance across measures. Users and hospitals should expect that differences in star ratings

across hospitals should be substantiated by clinically and statistically meaningful differences in underlying performance. As currently designed, star ratings continue to include measures with known methodological flaws (e.g., the patient safety indicator (PSI) composite measure). And concerns have been raised in the past about whether the LVM calculation was being executed correctly.

Stability. Any fluctuations in star ratings across reporting periods should be driven by significant changes in underlying measure performance rather than by any inherent instability in the ratings methodology. As advised by the AHA, CMS canceled the July 2018 update to star ratings in part because there were significant changes in star ratings. These rating changes were not explained easily by a major change in underlying measure performance.

A "line of sight" from star ratings to performance on underlying measures. Because star ratings are publicly reported, hospitals should be able to see how any positive or negative changes in underlying measure performance are reflected in their star ratings in a transparent and predictable fashion. Since their inception, hospitals have expressed frustration that they have virtually no way to predict how their performance on the underlying measures will translate into a star rating. This means the ratings are of little value to improvement efforts. In fact, they actually could discourage improvement efforts when hospitals work hard to improve an aspect of care and then see their star ratings go down.

Balanced assessment. Star ratings performance should be based on performance across the breadth of available measures, and not hinge disproportionately on only one or two measures. As noted in the public comment document, the PSI composite measure and hip/knee complication measure have historically dominated the score of the safety measure group, even though the infection measures likely represent higher priority issues.

Accounts for potential biases. The ratings must account adequately for differences in the clinical and social risk factors across the patients and communities that hospitals serve. Hospitals that serve sicker and poorer patients should be on a level playing field with all other hospitals. The AHA has repeatedly noted that the current approach to ratings disadvantages hospitals caring for poorer communities and those like academic medical centers that tend to care for higher complexity patient.

AHA'S ASSESSMENT OF CMS'S PROPOSED CHANGES

The attached table provides the AHA's assessment of the degree to which each star ratings change proposed by CMS would address the six design elements above. We would not expect that any single proposed change would address all of the "must have" elements and concerns that we have articulated. However, three of the suggested changes – empirical criteria for measure groups, peer grouping star ratings among similar hospitals, and using an "explicit" scoring approach not tied to the LVM – appear

to address partially at least three of these elements, and are worthy of further work by CMS. We comment briefly on each of these changes below.

New criteria for creating and maintaining measure groups. **The AHA supports CMS's proposed new clinical and empirical criteria for creating and maintaining star ratings measure groups.** CMS would use a three step approach – 1) an initial grouping based on clinical coherence, 2) a statistical "confirmatory factor analysis" that explores the extent to which there is a single factor that explains performance in the measure group; and 3) ongoing monitoring to ensure balance across the measures within the group.

We believe the confirmatory factor analysis would be especially helpful and important to implement. The fundamental premise of the LVM approach used in star ratings is that one can summarize the performance of the measures on an aspect of care (e.g., safety, mortality) into a single score that accounts for both actual performance and unobserved (or latent) performance. One way to test whether that assumption holds true is to use a confirmatory analysis to determine the extent of variation that is explained by the model. Performing this analysis on an ongoing basis would provide a stronger empirical basis for the measure groups, and identify groups that may need to be revised in the future.

Confidence interval-based measure weights. **The AHA agrees with CMS that the weights applied to the measures used in the LVM need to be revised. In particular, there is no reason to believe it is appropriate for the PSI-90 measures or the hospital-wide readmissions measure to be so disproportionately weighted in the calculation of star ratings such that they drown out the effect of other better – or at least equally good – measures in the safety and readmissions domains.** Based on the information available in the public comment document and communications we have had with experts in LVM, we believe the current approach "over-fits" the model and is methodologically wrong. We believe that by working with experts in LVM, it will be possible for CMS to develop a solution to this problem that is both mathematically correct and leads to a more rational approach for addressing measurement precision in star ratings, thereby improving the ratings accuracy, stability and balance.

In the star ratings LVM approach, CMS calculates a numerical "loading factor" for each star ratings measure. The higher a measure's loading factor, the more it drives performance within a particular measure group. As the AHA and others have repeatedly noted, the loading factors within the patient safety measure group have fluctuated significantly, even though performance on the underlying measures has not appreciably changed. Furthermore, two measures in particular – the PSI composite measure, and hip/knee complications – have a disproportionate influence on the safety score, even though the infection measures within the safety group arguably reflect more significant safety issues.

CMS asserts that at least some of the loading factor fluctuation and imbalance stem from the agency's approach to dealing with measure precision. CMS's current calculation of the LVM uses "denominator weights" in which hospitals are scored more heavily on measures that include larger numbers of patients. CMS offers three alternative approaches to this issue – confidence interval-based weights (in which the weights account for the confidence intervals of each measure's calculation), logarithm of the denominator-based weights and simply eliminating the denominator weights altogether.

CMS indicates that its preference would be to use a combination of current denominator weights and logarithm of the denominator weights. However, the data in the public comment document show that the confidence-interval based weights best improve the LVM model fit for the safety group, as well as the balance and stability of the safety measure group's loading factors. The AHA is concerned that continuing to use the current approach of denominator-based weights would only perpetuate the problems with star ratings.

Whatever other decisions are made about the calculation of the LVM, it first must be mathematically correct. We understand CMS and its contractor are trying to make it so, and we appreciate the staff's diligent efforts. There are LVM experts at many colleges and universities. We have shared the name of one such expert with CMS previously, and would be glad to provide additional names of experts and urge the agency to reach out to them.

Peer grouping. **The AHA believes CMS should continue to explore approaches to creating peer groups for star ratings as a short-term strategy to address the potential biases in star ratings. However, we also urge CMS to pursue further improvements to the risk adjustment approaches of its existing star ratings, as direct risk adjustment approaches may obviate the need for peer grouping in the future.**

To date, hospitals caring for sicker patients and poorer patients tend to fare worse on star ratings. Specifically, teaching hospitals, hospitals that report on larger numbers of star ratings measures, and hospitals receiving the highest disproportionate share hospital (DSH) payments (a proxy for the extent to hospitals serve the poor) all have ratings that are, on average, lower than other hospitals.

Peer grouping approaches attempt to create groupings of hospitals that are similar to one another on specific characteristics, comparing the performance of hospitals within those groupings. The basic notion is that it is fairer to compare hospitals that are similar to one another than it is to compare hospitals with very different characteristics. Furthermore, peer grouping is a viable approach to leveling the playing field in comparing hospital performance. Indeed, CMS already uses a peer grouping approach in its Hospital Readmissions Reduction Program (HRRP) by placing hospitals into peer

groups based on the proportion of dual-eligible patients they treat. This has resulted in some lowering of penalties for those caring for the poorest communities.

We believe CMS should continue to explore peer group stratification approaches as an interim step to improving the fairness of star ratings. The most promising variables to use in peer grouping should include those found to have an association to star ratings that are generally outside of the control of hospitals. These include the number of reported measures and the proportion of dual-eligible patients. CMS could consider peer groupings using only one of those two variables, or a peer grouping based on a composite of those two variables.

At the same time, we strongly urge CMS to view peer grouping as an interim strategy while it assesses ways to improve the risk adjustment of the measures in star ratings. As we have noted with CMS's implementation of dual-eligible peer grouping in the HRRP, there are some inherent shortcomings with peer grouping approaches. The use of peer groupings involves somewhat subjective choices about where to set the cut points of a particular group. For example, those hospitals at the upper end of one group and those at the lower end of the next group would have similar proportions of dual-eligible patients, but would be placed into different groups for performance comparison purposes. Furthermore, direct risk adjustment would help improve the precision of performance comparisons by ensuring that measure scores reflect the issues most relevant to each measured outcome. For example, in peer grouping, one has to assume that dual-eligible status is as large a determinant of performance for readmissions as it is for hip and knee complications, when in fact, the impact of dual-eligible status may be slightly different for each measure.

Explicit approach to star ratings. **The AHA believes a less complex "explicit" approach to scoring hospital star ratings may be the most promising long-term option for improving star ratings.** CMS's current approach to star ratings employs complex statistical modeling techniques (i.e., LVM, k-means clustering). We appreciate that CMS's intent in using these techniques was to create a rating that accounts for as many statistical vagaries as possible across the highly heterogeneous measures included in star ratings. Yet, as we noted above, the current methodology has led to an inaccurate and potentially biased picture of hospital quality. In addition, the use of such a statistically intensive methodology makes the ratings of virtually no use to hospital quality improvement efforts because it is nearly impossible for hospitals to predict how well they may perform on star ratings and the extent to which performance on any single measure drives their overall ratings.

CMS has indicated in the past that it views star ratings as tool for patients that was not intended to be used by hospitals to support quality improvement efforts. But the reality is that any data that are reported publicly can and do drive hospitals to seek to improve their performance or maintain a high level of performance. A star ratings approach with less uncertainty could help hospitals better benchmark their performance against others. Furthermore, hospitals are reporting that

private sector payers are increasingly expressing interest in using star ratings for contracting purposes. For these reasons, the continued use of a star ratings approach that is inherently unpredictable and not tied to hospital quality improvement efforts may no longer be tenable.

We encourage CMS to continue exploring a more explicit approach to star ratings. We acknowledge that a more explicit system would involve some choices about what measures to include, how to weight particular measures and what performance targets to set. But, CMS could consider adopting some more empirically-based approaches to assist in this work. For example, to identify the weights for particular groups of measures, CMS could undertake systematic surveying of patients to identify the aspects of quality that would be of the greatest importance to them. In addition, the criteria proposed in the public comment document for creating and maintaining measure groups could be adapted for use in a more explicit approach to star ratings.

Other proposed changes. As noted earlier, this letter's attachment includes the AHA's overall assessment of each of CMS's proposed changes. While we will not provide detailed comments on each of them, we note concerns with two proposals.

First, we strongly oppose any approach to scoring hospitals on individual components of the PSI composite measure in the safety measure group. In fact, the AHA continues to urge CMS to transition PSI measures out of all of its measurement programs. The AHA has long been concerned by the significant limitations of PSIs as a quality measure. PSIs use hospital claims data to identify patients who have potentially experienced a safety event. However, claims data cannot and do not fully reflect the details of a patient's history, course of care and clinical risk factors. As a result, the rates derived from the measures are highly inexact. PSI data may assist hospitals in identifying patients whose particular cases merit deeper investigation with the benefit of the full medical record. But, the measures are poorly suited to drawing meaningful conclusions about hospital performance on safety issues. In other words, PSIs may help hospitals determine what "haystack" to look in for potential safety issues. But the ability of the measure to consistently and accurately identify the "needle" (i.e., the safety event) is far too limited for use in public reporting and pay-for-performance applications. It is not surprising that a 2012 CMS-commissioned study showed that many of the individual components of PSI-90 have unacceptably low levels of reliability when applied to Medicare claims data.

Second, the AHA cannot support the "closed form" computational method to the LVM without further information about how this approach performs as compared to the current approach. CMS asserts that the closed form approach could produce results that are as accurate as the current "quadrature" approach while taking less time. But the agency does not provide any empirical information to support this assertion.

ADDITIONAL CONSIDERATIONS

While not included in the public comment proposal, the AHA believes CMS's ongoing work to improve hospital star ratings should address the three issues below. In fact, we would urge that CMS examine and address these issues before it implements any other changes to the star ratings.

Validation of computational approach. **The AHA urges CMS to engage a group of experts on LVM to ensure its calculation approach is executed correctly.** We greatly appreciated CMS's 2017 decision to suspend star ratings briefly and make some changes to how it was executing the existing methodology after discovering that some issues led to the misclassification of hospitals. Unfortunately, we believe there still may be problems leading to misclassification. This includes the need to correct the individual measure loading factors, but not by using confidence interval weightings as discussed above.

Readmissions measure outliers. **The AHA urges CMS to explore strategies to mitigate the impact of outliers in calculating the readmission measures used in star ratings.** A recent analysis from a team based at Rush University Medical Center showed that hospital performance on the readmission measure can be impacted dramatically by highly medically complex patients who require frequent re-hospitalizations. CMS could consider including additional exclusions in its readmission measure to ensure those hospitals caring for the most complex patients are not placed at an unfair disadvantage.

Topic-specific star ratings. **The AHA urges CMS to consider developing an alternative approach to star ratings in which, instead of an overall rating, hospitals receive ratings on specific clinical conditions or topic areas.** As we have noted in this letter, we believe there are ways in which CMS can improve its approach to creating a single overall star rating for hospital quality. At the same time, we continue to have significant concerns about the conceptual underpinnings of star ratings. The measures included in the ratings were never intended to create a single, representative score of hospital quality. Furthermore, the ratings often do not reflect the aspects of care most relevant to a particular patient's needs. For example, a family may be interested in selecting the best hospital for cancer care, but there are no such measures included in the current star ratings. That is why the AHA has encouraged CMS to consider developing an alternative approach in which star ratings are done by topic area such as patient safety, patient experience of care and cardiac care. This approach may lessen the possibility of consumers receiving misleading information about quality.

The AHA appreciates your consideration of these recommendations. We look forward to continuing to work with CMS to ensure star ratings achieve the goals of meaningfulness, accuracy and transparency that we and all stakeholders share. Please contact me if you have questions or feel free to have a member of your team contact Akin Demehin, director of policy, at ademehin@aha.org.

Sincerely,

/s/

Ashley Thompson
Senior Vice President
Public Policy and Policy Development

Attachment: AHA Assessment of How Proposed Star Ratings Changes Address "Must Have" Design Elements

		Useful to consumers	Accuracy	Stability	Line of sight to underlying measures	Balanced assessment	Accounts for potential bias	Pursue Further?
Proposed Short-term changes to existing methodology	Update star ratings once per year	Insufficient information	No	Partially	No	No	No	✗
	New empirical criteria for creating and monitoring measure groups	No	Partially	Partially	No	Partially	No	✓
	Divide safety measure group into medical and surgical safety subgroups	No	No	Partially	No	Partially	No	✗
	Use individual component PSI* measures	No	No	No	No	No	No	✗
	Confidence interval-based measure weights	No	No	No	No	Partially	No	✗
	Logarithm of denominator-based weights	No	Insufficient information	Insufficient information	No	Insufficient information	No	✗
	Eliminate denominator weights	No	No	No	No	Partially	No	✗
	Peer grouping	Partially	Partially	No	No	No	Partially	✓
	Weighted-average summary scores (e.g., combine two reporting periods)	No	No	Partially	No	No	No	✗
	Closed form computational method for latent variable model	Insufficient info	Insufficient information	Insufficient information	Insufficient information	Insufficient information	Insufficient information	✗
Proposed Long-term changes	Stop using LVM** and adopt an explicit approach to star ratings	Partially	Insufficient information	Partially	Yes	Partially	Insufficient information	✓
	Move away from K-means clustering	Insufficient information	Insufficient information	Insufficient information	Yes	No	No	✗
	Account for improvement over time	No	No	Partially	No	No	No	✗
	User-customized star ratings	Partially	No	No	No	Partially	No	✗

***PSI** = Patient Safety Indicator ****LVM** = latent variable model