

# **Summary of Technical Expert Panel (TEP) Evaluation of Hospital Quality Star Ratings on *Hospital Compare***

June 8, 2015

**Prepared by:**

Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation  
(YNHHSC/CORE)

This material was prepared by YNHHSC/CORE under contract to the Centers for Medicare & Medicaid Services (CMS), an agency of the U.S. Department of Health and Human Services. The contents presented do not necessarily reflect CMS policy.

## Table of Contents

Background .....	4
Star Ratings Development Team .....	4
The Technical Expert Panel .....	4
Specific Responsibilities of the TEP Members .....	5
TEP Members .....	6
TEP Meetings .....	7
First TEP Meeting .....	7
Second TEP Meeting .....	9
Third TEP Meeting.....	11
TEP Meeting Discussion and Feedback.....	14
Appendix A. Star Ratings Development Team .....	27
Appendix B. TEP Call Schedule .....	28
Appendix C. Pre-TEP Meeting 1 Survey .....	29
Appendix D. Phase 1 Measure Recommendations.....	30
Appendix E. Hospital Summary Score Modeling Approach.....	38
Simple average.....	38
Weighted average of individual measures .....	38
Latent variable models.....	38
Appendix F. Two-Stage Approach with Measure-Type Groups.....	40
Proposed Grouping Method .....	40
Proposed Approach: Two-Stage using Latent Variable Model and Weighted Averaging .....	41
Use of LVM in Two-Stage Approach .....	42
Appendix G. Proposed Weighting Options for Two-Stage Approach .....	43

Purposeful Weighting .....	43
Appendix H. Proposed Options for Star Ratings Translation .....	44
Option 1: Fixed Interval .....	44
Option 2: Overall and Measure-Specific Threshold .....	45
Option 3: k-Means Clustering .....	46
Appendix I. Approach for Missing Measures or Measure Groups.....	48
Minimum Number of Measures .....	48
Minimum Number of Domains .....	48
Recommendation for Minimum Thresholds for Public Reporting.....	48
Recommendation for Weighting .....	49
Appendix J. TEP Minutes .....	50

## Background

The Centers for Medicare & Medicaid Services (CMS) contracted with Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (CORE), Lantana Consulting Group, and the independent research institution NORC at the University of Chicago (NORC) to develop hospital quality star ratings on *Hospital Compare*. The purpose of the Overall Hospital Quality Star Ratings project is:

- To improve the usability and interpretability of *Hospital Compare* for patients and consumers; and
- To develop a methodology designed to generate an overall star rating for hospitals with sufficient quality data using the existing measures on *Hospital Compare*.

During development, the Star Ratings team has and will continue to obtain expert and stakeholder input on the Star Ratings methodology. The Star Ratings team meets regularly and is comprised of experts in policy and methodology. Additionally, the team convened a Technical Expert Panel (TEP) of 15 experts in consumer perspectives, purchaser perspectives, quality improvement, performance measurement, and healthcare disparities to provide input on key methodological decisions.

This report summarizes the feedback and recommendations provided by the TEP regarding the star rating system under consideration as of March 30, 2015.

## Star Ratings Development Team

The CORE Star Ratings development team is led by Dr. Arjun Venkatesh. The development team at Lantana is managed by Kit Cooper and Hector Cariello. The NORC development team is managed by Dr. Rachel Singer. (See [Appendix A](#))

## The Technical Expert Panel

In alignment with the CMS Measures Management System (MMS), the team released a 30-day public call for nominations and convened a TEP. The team contacted stakeholder groups and experts via email to acknowledge the public call for nominations.

The role of the TEP is to provide feedback on key methodological and analytical decisions made in consultation with the development team. The TEP is comprised of individuals with diverse perspectives and backgrounds and includes clinicians, researchers, consumers, purchasers, patients, and experts in quality improvement. The appointment term for the TEP is from December 2014 to September 2015.

## Specific Responsibilities of the TEP Members

- Review background materials provided by the Star Ratings team prior to each TEP meeting
- Participate in TEP conference calls
- Provide input on key clinical, methodological, and other decisions
- Provide feedback on key policy or other non-technical issues
- Review the TEP summary report prior to public release

## TEP Members

**Table 1. Members of the Star Ratings TEP**

Name	Organization	Location
Matt Austin, PhD	Johns Hopkins University School of Medicine (Assistant Professor)	Baltimore, MD
Vinita Bahl, DMD, MPP	Performance Assessment & Clinical Effectiveness, University of Michigan Health System (Director)	Ann Arbor, MI
John Bott, MSSW, MBA	State of Wisconsin Department of Employee Trust Funds (Manager of Performance Measurement) <i>Updated Affiliation (4/27/15):</i> Health Care Transparency/Performance Measurement, IPRO (Senior Director)	Madison, WI
Kathy Ciccone, RN, MBA	Healthcare Association of New York State Quality Institute (Executive Director)	Rensselaer, NY
Kelly Court, MBA	Wisconsin Hospital Association (Chief Quality Officer)	Madison, WI
Rachel Grob, PhD	The Center for Patient Partnerships, University of Wisconsin-Madison (Director of National Initiatives / Associate Clinical Professor)	Madison, WI
Rodney Hayward, MD	University of Michigan (Professor of Internal Medicine and Public Health)	Ann Arbor, MI
Emma Kopleff, MPH	Consumer-Purchaser Alliance, National Partnership for Women and Families (Senior Policy Advisor) <i>Updated Affiliation (3/18/15):</i> Community Health Accreditation Partner (CHAP) (Senior Standards and Research Analyst)	Washington, DC
Doris Peter, PhD	Consumer Reports Health Ratings Center (Director)	Yonkers, NY
Laura Petersen, MD, MPH, FACP	Michael E. DeBakey VA Medical Center (Associate Chief of Staff for Research)	Houston, TX
Casey Schwarz, JD	Medicare Rights Center (Policy & Client Services Counsel)	New York, NY
David Shahian, MD	Center for Quality and Safety, Massachusetts General Hospital (Vice-President)	Boston, MA
Brett Stauffer, MD, MHS	Baylor Scott and White Health (Director of Clinical Decision Support)	Dallas, TX
Guofen Yan, PhD	University of Virginia School of Medicine (Associate Professor of Biostatistics)	Charlottesville, VA
Ben Yandell, PhD	Clinical Information Analysis, Norton Healthcare (System Associate Vice President)	Louisville, KY

## TEP Meetings

The development team conducted the first meeting on December 01, 2014, the second meeting on February 4, 2015, and the third on March 30, 2015 ([Appendix B](#)).

- The first TEP meeting, summarized on pages 7 & 8, focused on selecting measures for inclusion in the Overall Hospital Quality Star Ratings.
- The second TEP meeting, summarized on pages 9 & 10, focused on the methodology for combining a hospital's available measure information into a single hospital summary score.
- The third TEP meeting, summarized on pages 10 & 11, focused on the methodology for translating hospital summary scores to star ratings.

### *First TEP Meeting*

During the first TEP meeting, the development team presented potential criteria to be used by CMS for measure inclusion or exclusion from the star ratings, along with high-level survey results provided by the TEP prior to the meeting ([Appendix C](#)).

For the purposes of this discussion, the TEP reviewed the list of measures reported on *Hospital Compare* as of October 2014.

Measures were selected based on the proposed criteria. A measure that did not fall into any of the criterion for exclusion or for further consideration was recommended for inclusion in the star ratings. Forty-seven of the 105 measures on *Hospital Compare* for October 2014 were recommended for inclusion. Thirty-three measures met one (or more) of the criterion for expert consideration and were brought to the TEP to weigh the advantages and disadvantages of including these measures in the star ratings.

The following bullets present a **high-level** summary of what was discussed during the first TEP meeting.

- Each TEP member was introduced and disclosed all relevant conflicts of interest.
- The development team began by presenting the fundamental aspects of the project including the project goals and timeline.
- The TEP approved the TEP charter.
- The development team then discussed the measure selection process for the star ratings.
  - Of the 105 measures on *Hospital Compare* as of October 2014, the developers recommended forty-seven measures for inclusion in the star ratings.
    - These forty-seven measures did not meet any of the exclusion criteria that were presented to the TEP.
  - The team recommended excluding twenty-five of the 105 measures according to the following criteria:
    - Measures retired, suspended, or delayed from public reporting on *Hospital Compare*;

- Ten measures did not have publicly available data on *Hospital Compare* for October 2014.
  - Publicly available data is required in order to ensure maximum transparency in the development and roll-out of a summary star rating.
- Measures awaiting public reporting for which hospital performance data is not currently publicly available on *Hospital Compare*; and
  - There are nine measures slated to be on *Hospital Compare* as of October 2014 that do not currently have performance data.
  - These measures are slated for future public reporting, therefore, should not be included in the current example star ratings.
- Measures for which less than 100 hospitals participate in public reporting.
  - There are six measures reported on *Hospital Compare* for which 100 or fewer hospitals report performance.
  - Including these measures without broad use will not support our goal to use widely available quality information to develop a valid summary star rating that is easily interpreted by patients and consumers.
- There are thirty-three measures that were discussed by the TEP for potential exclusion. These measures fell into four categories:
  - Measures “de-endorsed” by the National Quality Forum (NQF) and not supported by the Measure Applications Partnership (MAP);
  - Structural Measures;
    - Structural Measures (Volume)
    - Structural Measures (Yes/No)
  - “Topped Out” Measures; and
  - Efficiency Measures.

## Second TEP Meeting

During the second TEP meeting, the development team reviewed the measures selected for inclusion in the Overall Hospital Quality Star Ratings and the approaches under consideration for calculating hospitals' summary scores.

For the purposes of this discussion, the TEP reviewed preliminary analyses using December 2014 *Hospital Compare* data.

The development team had two main goals for the meeting:

1. Obtain feedback on the proposed approach for calculating hospital summary scores; and
2. Obtain feedback on two key decisions regarding the proposed modeling approach.

Below is a **high-level** summary of what was discussed during the second TEP meeting.

- The team began the second meeting by giving welcoming remarks and discussing the current status of the project. The project is structured around the following three phases of development:
  - Phase 1: Measure Selection;
  - Phase 2: Hospital Summary Scores; and
  - Phase 3: Overall Star Rating.
- The team reviewed the key decisions from Phase 1: Measure Selection, discussed during the first TEP meeting. The proposed measure exclusions for the star ratings project included:
  - Measures suspended, retired, delayed, or awaiting public reporting on *Hospital Compare*;
  - Measures with less than or equal to 100 hospitals reporting;
  - Structural measures without evidence of an association with changes in clinical practice or improved outcomes; and
  - Non-directional efficiency measures.

(See measure exclusion flowchart shown to the TEP is presented in [Appendix D](#)).
- The development team discussed the steps of the proposed hospital summary score methodology.
  - Measure Direction
    - For the 71 measures proposed for inclusion, the development team converted each measure to a “higher score is better” format.
    - This had no material impact on the analytic assumptions or ranking of hospitals.
  - Standardization
    - The development team discussed the need to standardize measure scores.
      - The 71 measures include a number of score formats including percentage, rate, ratio, and time.
      - In order to calculate a hospital summary score, the development team standardized each measure to have a mean score of 0 with a standard deviation of 1.
  - Winsorization
    - In order to prevent outliers from heavily impacting the summary score in the complex model, the development team used 99% Winsorization on extreme values.
    - Winsorization is an approach used to limit the effect of severe outlier hospitals on assessing hospital performance. In winsorization, hospitals with extremely high

(>99.5<sup>th</sup> percentile) or low performance (<0.5<sup>th</sup> percentile) on a measure are minimally adjusted to have a score at the 0.5<sup>th</sup> and the 99.5<sup>th</sup> percentile, separately. This approach improves the efficiency of calculations and the reliability of our methods without materially changing a hospital's measure performance score.

- Next, the development team must create a summary score. The team sought to identify an approach that would:
  - Generate a single, aggregate measure of available hospital quality information;
  - Account for the heterogeneity of measures available (process, outcome, etc.);
  - Account for the fact that different hospitals are reporting different numbers of measures;
  - Limit subjective assumptions;
  - Accommodate changes in the included measures and hospital performance over time; and
  - Utilize an evidence-based approach reflecting modern statistical methods that have been applied to healthcare research.
- With these principles in mind, the team proposed using a latent variable modeling approach. (See [Appendix E](#) for more details on the approaches considered).
- The TEP provided input on the decision to report an overall hospital star rating opposed to domain-level star ratings. Overall, the TEP advocated for an approach that accounted for the multidimensionality of the quality measures on *Hospital Compare*.
- The team concluded by discussing the next steps for the project.
  - Third TEP meeting in March 2015
  - Second public comment period in April 2015
  - Proposed dry run for June 2015

## Third TEP Meeting

After the second TEP meeting, the CORE team evaluated several modeling approaches based on feedback from the TEP. These approaches aimed to both capture the multidimensionality of quality data and satisfy the desired principles discussed during the second TEP meeting. Following review of these approaches within CMS, the team ultimately proposed a two-stage approach.

During the third TEP meeting, the development team sought to discuss several key elements of the proposed two-stage approach using measure-type groups to calculate hospital summary scores. The team also discussed the options for translating a hospital's summary score into a star rating. Lastly, the team sought feedback on the minimum measure and measure group thresholds required for a hospital to report an overall star rating.

The development team had four main goals for the meeting.

1. Review summary of comments received during first public comment period.
2. Review input on modeling and planned approach to generating hospital summary score.
3. Describe potential approach for translating summary score to star ratings.
4. Discuss options for handling missing measures and measure type groups.

For the purposes of this discussion, the TEP reviewed preliminary analyses using *Hospital Compare* data from December 2014 and April 2015.

The following bullets reflect a **high-level** summary of what was discussed during the third TEP meeting.

- The team began the third meeting by giving welcoming remarks and discussing the current status of the project.
  - Phase 1: Measure Selection
    - The team narrowed the list of included measures through public comment and expert input.
    - During the public comment, the team received 12 comments from stakeholders.
      - The TEP provided feedback on measures that are removed from public reporting outside of the rulemaking process. The team will use this feedback to improve its process for identifying measures excluded from the overall star ratings under the criterion, "measures suspended, retired, or delayed from public reporting on *Hospital Compare*."
    - The measures included in the star ratings as of April 2015 are listed in [Appendix D](#).
  - Phase 2: Hospital Summary Scores
    - The team proposed a two-stage approach to calculate hospital summary scores in which measures are categorized into measure groups for initial modeling, and group scores are subsequently aggregated using purposeful (conceptual) weights. (For more information on the proposed Hospital Summary Score Approach, please see [Appendix F](#) and [Appendix G](#)).
  - Phase 3: Overall Star Ratings
    - The team presented three potential approaches for categorizing hospitals' summary scores into star ratings.
  - Both Phase 2 and Phase 3 will be presented for public comment.

- The team described the comments received during public comment for the first phase of the project.
  - The summary report from the first public comment period will be publicly posted by Spring 2015.
- Next, the team reminded the TEP of the goals for the modeling approach.
  - Produce a single hospital summary score.
  - Capture the multidimensionality of data.
  - Account for any missing measures or measure groups for each hospital.
  - Replicate every quarter with measures being added or removed.
  - Allow for policy-driven weighting.

### **Hospital Summary Scores**

- The team presented the proposed two-stage approach for Phase 2: Hospital Summary Scores.
  - The First Stage: 6 separate latent variable models to reflect quality of hospital care along separate measure groups
  - Six measure-type groups were proposed for this calculation.
    - Patient Experience Measures (HCAHPS)
    - Outcome: Safety Measures
    - Outcome: Readmission Measures
    - Outcome: Mortality Measures
    - Process Measures
    - Efficiency Measures
  - The Second Stage: Weighted average of a hospital's available measure group scores
  - More information about the proposed approach can be found in [Appendix F](#).

### **Weighting and Measure-Type Groups**

- Next, the team discussed weighting options for the proposed two-stage approach, considering two choices:
  - Equal weighting for all measure-type groups; and
  - Modified weighting based on the Hospital Based Purchasing (HVBPP) Program weighting.
- Prior to the TEP, the team distributed a survey to understand measure group importance.
  - Qualitative comments demonstrated that mortality and safety groups were broadly viewed as most important relative to the other groups.
- During the TEP discussion, some TEP members felt the process measures should be weighted highest.

### **Star Ratings Translation**

- The development team presented three options for translating hospital summary scores into star ratings ([Appendix H](#)).
  - Option 1: Fixed Intervals
    - This would classify hospitals within certain percentiles of the star ratings.
  - Option 2: Overall and Measure-Specific Threshold
    - This would classify hospitals based on their summary and measure score compared to the national averages.
  - Option 3: *k*-Means Clustering
    - This would classify to be based on statistical clustering around five points.

- Next, the team discussed how to account for missing measures or measure-type groups ([Appendix I](#)).
- Out of the 4,753 hospitals with information on *Hospital Compare* as of December 2014, the mean number of measures per hospital was 42 measures.
- The team proposed the idea of creating a minimum threshold for both the number of measures and the number of measure-type groups needed to calculate a summary score.

#### **Next Steps**

- Next steps for this project will include:
  - Second public comment period in the spring 2015; and
  - Dry run proposed for June 2015.

## TEP Meeting Discussion and Feedback

**Table 2. Key Issues Discussed During First TEP Meeting and Feedback**

Key Issues Discussed	TEP Feedback/Discussion
Star ratings maintenance process	<p>One TEP member asked about future review of these measure categories on an ongoing maintenance schedule.</p> <p>The development team responded that the approach to maintenance of star ratings has not been fully laid out. However, the goal is to have an approach for adding/removing measures and refreshing data each quarter. The criteria developed by the TEP will help to guide these efforts.</p>
NQF de-endorsed and MAP measures	<p>One TEP member asked if measures that are “de-endorsed” by NQF and not supported by the MAP would be likely to be removed from the <i>Hospital Compare</i> website.</p> <p>The team responded that, historically, such measures have commonly been retired from <i>Hospital Compare</i>. The discussion lead to the recommendation that this criteria be removed (i.e., such measures would be included in star ratings while they remain on <i>Hospital Compare</i>) and, therefore, star ratings measures would align with CMS policy and what is available on <i>Hospital Compare</i>.</p> <p>Some TEP members felt NQF de-endorsement was a strong statement that may be meaningful to patients and consumers.</p> <p>The TEP generally disapproved of the criteria to exclude measures “de-endorsed” by NQF and not supported by the MAP. The TEP would like the star ratings to be harmonized with CMS’s decisions for public reporting. Overall, the TEP did not reach a consensus, but provided input for CMS to finalize this decision.</p>
Measurement gaps	<p>One TEP member asked if they could provide recommendations for results they would like to see added to the website.</p> <p>The team responded that recommendations for measurement gaps would be reasonable, and we would welcome individual feedback.</p>

Key Issues Discussed	TEP Feedback/Discussion
Population being measured	<p>One TEP member expressed concern regarding selecting measures aimed at a specific population that uses the star ratings on <i>Hospital Compare</i> most frequently.</p> <p>CORE expressed that audience for star ratings would include all age groups and populations. Many of the measures use data from patients 65 years or older, however, the site is a public site that can be used by any individual.</p>
Structural Measures	<p>One TEP member noted that structural measures do help inform quality, but asked for clarification on whether decisions to include or exclude this type of measure would be continued on through maintenance of the star ratings system.</p> <p>One TEP member offered support to exclude the specific structural dichotomous measures that were included in this discussion, while also agreeing that if there are changes in measure methodology or measures added to this measure type, the team may want to reconsider this criterion.</p> <p>One TEP member expressed support for not including the OP-17 and OP-12 measures, stating that HIT measures are not informative structural measures for patients. This member also added that all hospitals are going to have electronic health records (EHRs), but at this point in time, the member is still skeptical of the OP-12 measure being helpful for patients while different platforms are being used.</p> <p>Next, the group discussed the surgery check list measure (OP-25). One TEP member did not support including this measure in the star ratings methodology as the measure is “yes” or “no,” instead of specifying if a surgery was done in an appropriate manner.</p> <p>Another TEP member supported that statement and added that there was a recent study outlining the variability of check list implementation in hospitals.</p> <p>Another TEP member supported these statements saying that this would be considered a “low-bar measure” by some groups.</p>

Key Issues Discussed	TEP Feedback/Discussion
	<p>One TEP member felt that having volume measures available to consumers was very important. However, they did not feel it was needed for the star ratings.</p> <p>Another TEP member agreed with the comments saying that based on the list of outpatient procedures that are included, they would not include volume measures in the star ratings.</p> <p>Another TEP member felt that there is a wide range of performance on these measures, so it is important for hospitals to see these volume results; however, they did not feel this added to the overall quality.</p> <p>Another TEP member felt that for the consumer, there needed to be a compelling link to an outcome for volume measures to be meaningful. They added that patients are possibly more concerned with volume at the surgeon-level than at the hospital-level.</p> <p>The development team responded that the feedback seemed to point to not including these particular measures. However, conceptually, volume measures are important and should be reconsidered as other measures are up for consideration, particularly for conditions or procedures with strong evidence of a volume-outcome relationship.</p> <p>The TEP generally supported exclusion of registry measures, the current structural measures of volume, and other dichotomous structural measures without evidence of an association with changes in clinical practice or improved outcomes. Structural measures of volume may be useful if the procedural volume has substantial evidence of an association with improved outcomes. In other words, inclusion of this category of measures is dependent on the procedures included in the measure and the supporting literature.</p>
Topped out measures	<p>One TEP member added support for including “topped out” measures, stating that if hospitals have good performance, they should get credit for it. This member added that if the measures continue to be active on <i>Hospital Compare</i>, then they should be included in the star ratings.</p>

Key Issues Discussed	TEP Feedback/Discussion
	<p>Another TEP member expressed concern that we would end up with many hospitals that are five star hospitals with this information. This would lead to consumers having difficulty differentiating performance.</p> <p>The developers responded that there are two different ideas of “topped out,” one being topped out in star performance, while the other is topped out measure performance. If we recommended to include these measures, a future discussion will be needed to determine how to incorporate these measures into the star ratings, and would not necessarily mean that the stars would be “topped out.”</p> <p>Another TEP member felt it would best to exclude these measures stating that when performance is extremely compressed, it is not reliable for discriminating among performance levels.</p> <p>Another TEP member felt that these measures were reliable, but are topped out because of their use in programs. They added that the fact that there are outliers in these measures does have real meaning for quality.</p> <p>Another TEP member added that when considering measures geared towards consumers, they would not consider including measures like these.</p> <p>The TEP did not reach a consensus on this measure category. CORE will take the TEP’s feedback to CMS for further discussion.</p>
Efficiency Measures	<p>One TEP member recommended having a separate star for efficiency measures because efficiency seems like a counter balance to quality and is an important part of what is driving the healthcare industry.</p> <p>Another TEP member expressed concern with the exclusion of these measures as they could serve as a proxy for the risk of harm from overuse.</p> <p>One TEP member added that the measures were built off of coding which may not reflect true underlying clinical risk.</p>

Key Issues Discussed	TEP Feedback/Discussion
	<p>CORE responded that the TEP discussion should focus on use of the measures in a rating system given the known limitations and strengths of each measure type.</p> <p>Several TEP members acknowledged the importance of efficiency measures within the context of the Triple Aim and supported inclusion in star ratings, while some TEP members questioned the relationship between efficiency measures and quality.</p> <p>Other TEP members discussed the relationship between imaging overuse and many aspects of quality such as radiation safety and unnecessary downstream testing. TEP members agreed that non-directional efficiency measures for which a higher or lower score is not necessarily “better” should not be included in star ratings. Regarding directional measures, the development team will take the TEP’s feedback to CMS for further discussion.</p>

**Table 3. Key Issues Discussed During Second TEP Meeting and Feedback**

Key Issues Discussed	TEP Feedback/Discussion
Feedback on Phase 1: Measure Selection	<p>One TEP member asked if the “topped-out” measure discussed are mentioned in the latest Inpatient Prospective Payment System (IPPS) rule.</p> <p>CORE responded that the discussed measures reflect the most recent published IPPS rule as well as MAP and NQF documentation. CORE added that if a measure was recommended for removal by the IPPS rule, it would be excluded based on that criterion.</p> <p>CMS added that the IPPS and Outpatient Prospective Payment System (OPPS) rules for FY 2017 are in the process of being drafted and have not yet been determined for the upcoming year.</p> <p>Another TEP member recommended that the team uses flexible criteria for inclusion as measurement improves and changes.</p>

Key Issues Discussed	TEP Feedback/Discussion
	CORE responded that the goal is to create a flexible methodology, approach, and criteria that could change as new measures become available.
Modeling Approach	<p>A TEP member mentioned that the latent variable analysis would reverse measures with a negative score naturally. The TEP member added that this additional step, while unnecessary, will not create any problems for calculating summary scores.</p> <p>One TEP member was concerned that latent variable analysis may reward hospitals with good performance on process measures because process measures dominate the model. The TEP member argued that if this is the case, a hospital's summary score might not be indicative of their performance on outcome measures.</p> <p>CORE explained that the loadings are not weights for each measure or measure-type grouping. Loadings represent the summary score's relationship with a given measure. The higher loadings do not alone indicate that a hospital can have a high summary score.</p> <p>Another TEP member asked if there had been any consideration of incorporating confidence intervals into the summary score.</p> <p>CORE confirmed that the team has considered incorporating confidence intervals into its presentation of the summary scores to hospitals and for informing the assignment of star ratings.</p>
Latent Variable Analyses: Shrinkage	<p>One TEP member stated that shrinkage will have little effect on the model because the model aggregates measures. The TEP member added that shrinkage is of greater concern at the individual measure level.</p> <p>One TEP member asked if it is possible to do a missing data imputation to reduce shrinkage before applying a modeling approach.</p> <p>CORE responded that previous feedback suggested that stakeholders do not like imputation and hospitals feel as though they are being assigned a score that lacks a meaningful application to their hospital.</p>

Key Issues Discussed	TEP Feedback/Discussion
Latent Variable Analyses: Quality Dimensions	<p>One TEP member asked if it is possible that more than one dimension of quality affects the summary scores when discussing the latent quality trait.</p> <p>CORE added that the current modeling approach assumes quality can be measured as a single dimension. CORE encouraged the TEP to give feedback on the Latent variable model as there were conceptual and analytic challenges expected when measuring quality as having a single or multiple dimensions.</p> <p>Another TEP member stated that latent variable analysis is designed to test if there is a single dimension or multiple dimensions of quality.</p>
Latent Variable Analyses: Loadings	<p>One TEP member mentioned that lower loadings may also be very important and that the number of measures that reflect the same concept will influence hospitals' summary scores.</p> <p>CORE recognized that the number of measures reflecting the same concept will influence hospitals' summary score. This approach models a single underlying quality trait, accounts for the relationship between measures by predicting random-effect, and empirically derives loadings flexible to changes over time.</p> <p>One TEP member asked if the loadings were unique to a hospital.</p> <p>CORE responded that loadings are unique to each measure, not each hospital.</p> <p>One TEP member did not agree with keeping measures with negative loadings because it penalizes those hospitals that are actually doing better and suggested dropping these measures if the variation is minimal.</p>
HCAHPS	<p>One TEP member asked where the 11 HCAHPS measures come from and if CORE has considered including some number of HCAHPS measures between 11 and one in the model.</p> <p>CMS added that HCAHPS publicly reports on 11 HCAHPS measures, some of which are composites, and the HCAHPS</p>

Key Issues Discussed	TEP Feedback/Discussion
	star rating will be publicly reported in April <i>on Hospital Compare</i> .
Overall Star vs. Domain Star	<p>One TEP member asked for clarification on the need for a single overall star rating versus domain-level star ratings.</p> <p>CMS responded that at this time, the project goal is to create a summary star rating for all hospitals. CMS added that domain-level star ratings would be addressed in the future.</p> <p>Several TEP members stated a preference for starting with the overall score and working towards the eventual ability to drill down into the domains. Several TEP members stated the need to respect what the consumers have identified as being most useful or important to them.</p> <p>Another TEP member expressed concern that aggregating measures that were empirically different would not actually produce a useful and meaningful measure for the consumer.</p> <p>CORE responded that initially developing this model, they tried to allow the data to define several latent variables through a form of factor analysis. The model did not converge using factor analysis because of missing data and the variability in the number of available measures across hospitals.</p> <p>CORE added that in order to develop domains, there would need to be conceptual definitions and would require a process for measure assignment.</p>

**Table 4. Key Issues Discussed During Third TEP Meeting and Feedback**

Key Issues Discussed	TEP Feedback/Discussion
Overall Star Ratings	<p>The TEP discussion began with a few comments regarding the overall project objectives and acknowledgement of the wide variety of stakeholder opinions and concerns regarding underlying measure validity and the relative weighting of measures.</p> <p>In response to one TEP comment regarding the limited validity of measures that are not adjusted for socio-</p>

Key Issues Discussed	TEP Feedback/Discussion
	demographic status (SDS), CORE noted that SDS is an active area of discussion with CMS and NQF, and as the measure methodologies evolve, the star ratings methodology will evolve as well.
Two-stage Approach: Measure-group types and loadings	<p>One TEP member suggested that since the loadings would change every three months, the documentation of measure loadings should also be updated each time.</p> <p>CORE responded that the team is still planning the update and maintenance processes. CORE asked if the TEP felt that the loadings should be publicly posted each quarter. CORE added that while the loadings might not be identical, the team does not anticipate them changing dramatically from one reporting period to the next.</p> <p>One TEP member recommended updating and posting the loadings to promote maximum transparency and help hospitals and consumers better understand the star ratings.</p> <p>Another TEP member asked what the process would be for adding new measure groups.</p> <p>CORE responded that as measures are added or removed, the process for maintaining these groups may need to change. These measure-type groups were created based on the existing measures on <i>Hospital Compare</i>. The team will look to CMS for guidance on how best to add/remove measure groups.</p> <p>One TEP member asked for clarification on the initial grouping of the process measure group, asking if the team had looked at separating these measures by outpatient and inpatient measures.</p> <p>CORE responded that this is continuing work, but the team will consider the proposed idea for grouping process measures.</p> <p>Another TEP member asked what would be done with the measures that have been discontinued. The member added that these measures are not being actively collected, but are still displayed on <i>Hospital Compare</i>.</p>

Key Issues Discussed	TEP Feedback/Discussion
	<p>CORE responded that they would follow-up for more information on these measures and take this to CMS.</p>
Two Stage Approach: Weighting	<p>One TEP member said the measure groups may not make sense with the signal of quality the measures display. This TEP member added that patient experience measures should be a star rating outside of the overall star rating so that there is a patient satisfaction rating and technical rating.</p> <p>One TEP member stated that the star ratings project could be gamed due to the type of measures included in the project. This member added that the process and patient satisfaction measures are harder to game, but felt that outcome measures lacked the risk adjustment needed for this project. The member expressed concern for giving these measure groups a high weighting in star ratings.</p> <p>Another TEP member added that the weighting could be determined based on the number of measures in the domain. The member added that there are issues with the risk adjustment for the PSI-4 measures as well.</p> <p>CORE responded that they appreciate the wide range of input. They added that knowing that measures evolve, the methodology for star ratings may evolve as well.</p> <p>One TEP member said it will create mixed signals and confusions if this doesn't align with other CMS programs. The TEP member added that if a hospital got penalized for poor quality but received a high star rating, this will create confusion for stakeholders.</p> <p>One TEP member supported the modified HVBP weighting and stated the weights achieved a fair balance among measure groups. The TEP member added that this process will not be a "black box" for hospitals if hospitals are able to recreate their scores.</p> <p>CORE responded that hospitals' star ratings will be a part of a preview period, and the calculation of the star rating will be made as transparent as possible. CORE also aims to be as clear as possible in technical documents and with the display on <i>Hospital Compare</i>.</p>

Key Issues Discussed	TEP Feedback/Discussion
	<p>Another TEP member added that the intent is to display what consumers might want to see and that the literature shows consumers like outcomes measures. The TEP member added that while there may be weaknesses in certain measures, it is important to not redo the work of the National Quality Forum (NQF).</p> <p>One TEP member expressed concerns about the weight for the readmission measures as they felt that readmission performance is dependent on the hospital's location and patient population. The TEP member recommended having a lower weighting for the readmission group.</p> <p>Another TEP member generally supported the outcomes groups and noted that the risk models are increasingly improving. This member supported the proposed weighting for the process measures.</p> <p>Another TEP member recommended grounding the weighting in a criteria-based process so that adding or removing measure groups may be less difficult.</p> <p>CORE responded that they are open to other options for adding criteria for weighting and will bring recommendations back to CMS.</p>
Star Ratings Translation	<p>One TEP member asked for some clarification regarding what approaches other star ratings efforts are using.</p> <p>William Lehrman, CMS added that HCAHPS is considering <i>k</i>-means clustering and that this method is used for other star ratings.</p> <p>Another TEP member recommended using the <i>k</i>-means clustering option as the frequencies generated by this option feel more like what would be expected.</p> <p>One TEP member said option 2 or option 3 would be closer to a consumer interpretation of star frequencies. The TEP member recommended the rationale for the selecting an option be made publicly available.</p> <p>Another TEP member added that option 2 or option 3 are both defensible. The TEP member added that regardless of what approach is used, there will be two adjacent hospitals</p>

Key Issues Discussed	TEP Feedback/Discussion
	<p>that have just different enough scores that will be in different categories.</p> <p>Another TEP member asked for clarification on how option 2 or option 3 differ in classification.</p> <p>Another TEP member added that they would favor option 2 or option 3 over option 1. The TEP member added that star ratings should incentivize hospitals to do well and not cap how well hospitals can do.</p> <p>Another TEP member added that option 1 was, on the surface, appealing philosophically because it distributes hospitals more evenly across each of the five star categories. This member added that option 2 is concerning, and if an overwhelming majority of hospitals receive a three-star rating, then the ratings won't be as helpful to patients and consumers.</p>
Missing Measures and Missing Measure Groups	<p>One TEP member expressed concern with programs that transfer all weight to the groups a hospital has available. The TEP member added that a hospital should be required to have a minimum number of measures in each group.</p> <p>One TEP member suggested that if a hospital was missing an important group, such as a highly weighted group like Safety, the hospital would not receive a star rating. The TEP member added that if the hospital was missing a lower weighted domain, the weight could be redistributed.</p> <p>CORE responded that they recognized the challenges of redistribution and will take the recommendations back to CMS.</p> <p>Another TEP member recommended instead of using a fixed number of measures, CORE could apply a percentage threshold where a hospital must report a certain percentage of measures in a group to receive a star rating.</p> <p>CORE responded that this could be a reasonable approach and will present this option to CMS.</p> <p>One TEP member recommended that CMS display the measure-type group summary scores instead of just the overall rating for the hospital.</p>

Key Issues Discussed	TEP Feedback/Discussion
	<p>CORE said that this feedback will be presented to CMS; however, the scope of this work is focused on developing a methodology for an overall star rating at this time.</p> <p>Another TEP member stated that the outcome measure groups are the smallest in terms of the number of measures per group. The TEP member advocated for creating broader groups for the measures.</p> <p>CORE thanked the TEP for the recommendations.</p>

## Appendix A. Star Ratings Development Team

**Table 5. CORE Members of the Star Ratings Development Team**

Name	Title/Affiliation	Contact Information
Arjun Venkatesh, MD, MBA, MHS	Project Lead	<a href="mailto:arjun.venkatesh@yale.edu">arjun.venkatesh@yale.edu</a>
Susannah Bernheim, MD, MHS	Project Director	<a href="mailto:susannah.bernheim@yale.edu">susannah.bernheim@yale.edu</a>
Jaymie Potteiger, MPH	Project Coordinator	<a href="mailto:jaymie.potteiger@yale.edu">jaymie.potteiger@yale.edu</a>
Angela Hsieh, PhD	Lead Analyst	<a href="mailto:angela.hsieh@yale.edu">angela.hsieh@yale.edu</a>
Haiqun Lin, MD, PhD	Statistical Consultant	<a href="mailto:haiqun.lin@yale.edu">haiqun.lin@yale.edu</a>
Jeph Herrin, PhD	Analyst	<a href="mailto:jeph.herrin@yale.edu">jeph.herrin@yale.edu</a>
Harlan Krumholz, MD, SM	Principal Investigator	<a href="mailto:harlan.krumholz@yale.edu">harlan.krumholz@yale.edu</a>
Zhenqiu Lin, PhD	Analytics Manager	<a href="mailto:zhenqui.lin@yale.edu">zhenqui.lin@yale.edu</a>
Benjamin Clopper, MPH	Research Associate	<a href="mailto:benjamin.clopper@yale.edu">benjamin.clopper@yale.edu</a>
Mallory Perez, BSPH	Research Assistant II	<a href="mailto:mallory.perez@yale.edu">mallory.perez@yale.edu</a>
Erica Norton, BS	Research Assistant II	<a href="mailto:erica.norton@yale.edu">erica.norton@yale.edu</a>

**Table 6. Lantana Consulting Group Members of the Star Ratings Development Team**

Name	Title/Affiliation	Contact Information
Kit Cooper, BS	Lantana Project Manager	<a href="mailto:kit.cooper@lantanagroup.com">kit.cooper@lantanagroup.com</a>
Hector Cariello, MPH	Lantana Project Manager & Program Analyst	<a href="mailto:hector.cariello@lantanagroup.com">hector.cariello@lantanagroup.com</a>

**Table 7. NORC Members of the Star Ratings Development Team**

Name	Title/Affiliation	Contact Information
Romonda Bumpus	NORC Research Scientist	<a href="mailto:Bumpus-Romonda@norc.org">Bumpus-Romonda@norc.org</a>
Alexis Estomin, BA	NORC Research Assistant	<a href="mailto:Estomin-Alexis@norc.org">Estomin-Alexis@norc.org</a>
Rachel Singer, PhD, MPH, MPA	NORC Senior Research Scientist	<a href="mailto:Singer-Rachel@norc.org">Singer-Rachel@norc.org</a>

## Appendix B. TEP Call Schedule

1. December 1, 2014 – 5:00pm – 7:00pm EST
2. February 4, 2015 – 5:00pm – 7:00 pm EST
3. March 30, 2015 – 4:00-6:00 pm EST

## Appendix C. Pre-TEP Meeting 1 Survey

Prior to the TEP meeting, the development team distributed an online survey for each of the categories and specific measures within each category.

The members were allowed to select an overall category response or to rate each individual measure within a category for inclusion/exclusion.

- Measures “de-endorsed” by NQF and not supported by MAP:
  - 36% of TEP members supported inclusion of these measures
  - 64% of TEP members supported exclusion or selection by measure
- Structural Measures:
  - Structural measures with dichotomous outcomes
    - 29% of TEP members supported inclusion of these measures
    - 71% of TEP members supported exclusion or selection by measure
  - Structural measures of procedural volume
    - 50% of TEP members supported inclusion of these measures
    - 50% of TEP members supported exclusion or selection by measure
- “Topped out” Measures:
  - 43% of TEP members supported inclusion of these measures
  - 57% of TEP members supported exclusion or selection by measure
- Efficiency Measures:
  - 29% of TEP supported inclusion of these measures
  - 71% of TEP members supported exclusion or selection by measure

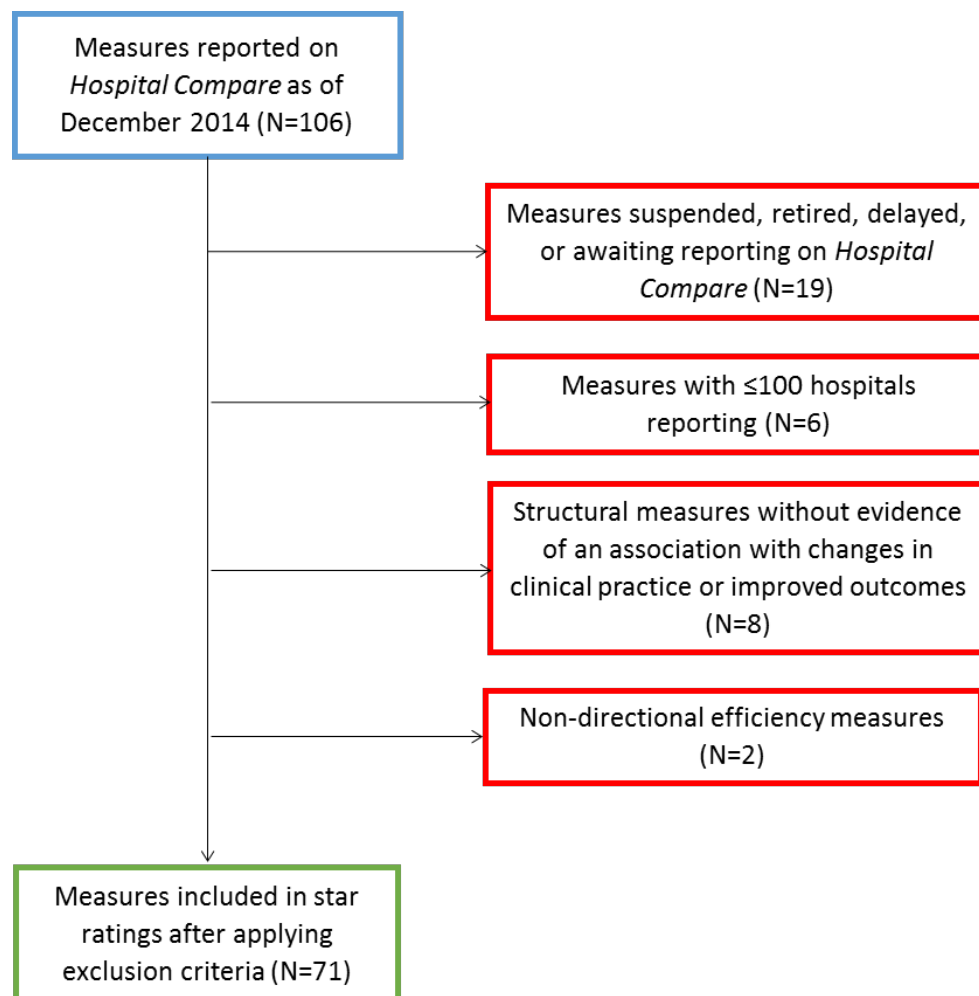
## Appendix D. Phase 1 Measure Recommendations

**Table 8. Description of measures included in overall star ratings (October 2014 data)**

Measure Characteristic	Number of Measures (N=71)
Measure Type	
Process	37
Outcome	18
Efficiency	5
Patient Engagement	11
NQF Endorsement Status	
Endorsed	58
De-endorsed	8
Not Endorsed	2
Endorsed – Time Limited	2
Endorsed Reserve	1
Primary Data Source	
Medical Record	39
Claims	16
National Healthcare Safety Network (NHSN)	5
HCAHPS	11

Note: October 2014 was presented to the TEP during the first TEP meeting when the characteristics of measures on *Hospital Compare* were discussed.

**Figure 1. Measure selection flowchart (December 2014 data)**



Note: This figure reflects key measure selection decisions as of TEP Meeting 2. Additional measures have been added/removed from *Hospital Compare* as of April 2015 (see Table 9 and Table 10).

**Table 9. Measures included in overall star ratings (April 2015 data)**

Measure Group	Measure Name (N=75)
Outcome-Mortality	<b><u>MORT-30-AMI</u></b> Acute Myocardial Infarction (AMI) 30-Day Mortality Rate
Outcome-Mortality	<b><u>MORT-30-HF</u></b> Heart Failure (HF) 30-Day Mortality Rate
Outcome-Mortality	<b><u>MORT-30-PN</u></b> Pneumonia (PN) 30-Day Mortality Rate
Outcome-Mortality	<b><u>MORT-30-COPD</u></b> Chronic Obstructive Pulmonary Disease (COPD) 30-day Mortality Rate
Outcome-Mortality	<b><u>MORT-30-STROKE</u></b> Stroke 30-day Mortality Rate
Outcome-Mortality	<b><u>PSI-4-SURG-COMP</u></b> Death Among Surgical Patients with Serious Treatable Complications (Harmonized with Nursing Sensitive Care Measure, Failure to Rescue)
Outcome-Safety	<b><u>COMP-HIP-KNEE</u></b> Hip/Knee Complications Hospital-Level Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and Total Knee Arthroplasty (TKA)
Outcome-Safety	<b><u>HAI-1</u></b> Central-Line Associated Bloodstream Infection (CLABSI)
Outcome-Safety	<b><u>HAI-2</u></b> Catheter-Associated Urinary Tract Infection (CAUTI)
Outcome-Safety	<b><u>HAI-3</u></b> Surgical Site Infection from colon surgery (SSI-colon)
Outcome-Safety	<b><u>HAI-4</u></b> Surgical Site Infection from abdominal hysterectomy (SSI-abdominal hysterectomy)
Outcome-Safety	<b><u>HAI-5</u></b> MRSA Bacteremia
Outcome-Safety	<b><u>HAI-6</u></b> <i>Clostridium Difficile</i> ( <i>C.difficile</i> )
Outcome-Safety	<b><u>PSI-90-Safety</u></b> Complication/Patient Safety for Selected Indicators (PSI) (composite)
Outcome-Readmission	<b><u>READM-30-AMI</u></b> Acute Myocardial Infarction (AMI) 30-Day Readmission Rate
Outcome-Readmission	<b><u>READM-30-HF</u></b> Heart Failure (HF) 30-Day Readmission Rate
Outcome-Readmission	<b><u>READM-30-Hip-Knee</u></b> Hip/Knee Readmission Hospital-Level 30-Day All-Cause Risk- Standardized Readmission Rate (RSRR) Following Elective Total Hip Arthroplasty (THA)/Total Knee Arthroplasty (TKA)
Outcome-Readmission	<b><u>READM-30-HOSP-WIDE</u></b> Hospital-Wide All-Cause Unplanned Readmission (HWR)
Outcome-Readmission	<b><u>READM-30-PN</u></b> Pneumonia (PN) 30-Day Readmission Rate
Outcome-Readmission	<b><u>READM-30-COPD</u></b> Chronic Obstructive Pulmonary Disease (COPD) 30-day Readmission Rate
Outcome-Readmission	<b><u>READM -30-STROKE</u></b> Stroke 30-day Readmission Rate

Measure Group	Measure Name (N=75)
Patient Experience	<b><u>H-HSP-RARTING</u></b> Overall Rating of Hospital (Q21)--Global Items
Patient Experience	<b><u>H-CLEAN-HSP</u></b> Cleanliness of Hospital Environment (Q8)--Individual Items
Patient Experience	<b><u>H-COMP-1</u></b> Nurse Communication (Q1, Q2, Q3)--Composite Topics
Patient Experience	<b><u>H-COMP-2</u></b> Doctor Communication (Q5, Q6, Q7)--Composite Topics
Patient Experience	<b><u>H-COMP-3</u></b> Responsiveness of Hospital Staff (Q4, Q11)--Composite Topics
Patient Experience	<b><u>H-COMP-4</u></b> Pain management (Q13, Q14)--Composite Topics
Patient Experience	<b><u>H-COMP-5</u></b> Communication About Medicines (Q16, Q17)--Composite Topics
Patient Experience	<b><u>H-COMP-6</u></b> Discharge Information (Q19, Q20)--Composite Topics
Patient Experience	<b><u>H-QUIET-HSP</u></b> Quietness of Hospital Environment (Q9)--Individual Items
Patient Experience	<b><u>H-RECMND</u></b> Willingness to Recommend Hospital (Q22)--Global items
Patient Experience	<b><u>H-COMP-7</u></b> HCAHPS 3 Item Care Transition Measure (CTM-3)
Process	<b><u>AMI-8a</u></b> Timing of Receipt of Primary Percutaneous Coronary Intervention (PCI)
Process	<b><u>ED-1b</u></b> Median Time from Emergency Department (ED) Arrival to ED Departure for Admitted ED Patients – Reporting Measure
Process	<b><u>ED-2b</u></b> Admit Decision Time to ED Departure Time for Admitted Patients - Reporting Measure
Process	<b><u>HF-2</u></b> Evaluation of LVS Function
Process	<b><u>IMM-2</u></b> Influenza Immunization
Process	<b><u>IMM-3-FAC-ADHPCT</u></b> Healthcare workers given influenza vaccination
Process	<b><u>OP-3</u></b> Median Time to Transfer to Another Facility for Acute Coronary Intervention
Process	<b><u>OP-4</u></b> Aspirin at Arrival
Process	<b><u>OP-5</u></b> Median Time to ECG
Process	<b><u>OP-6</u></b> Timing of Antibiotic Prophylaxis
Process	<b><u>OP-7</u></b> Prophylactic Antibiotic Selection for Surgical Patients
Process	<b><u>OP-18b/ED-3</u></b> Median Time from ED Arrival to ED Departure for Discharged ED Patients
Process	<b><u>OP-20</u></b> Door to Diagnostic Evaluation by a Qualified Medical Professional
Process	<b><u>OP-21</u></b> ED-Median Time to Pain Management for Long Bone Fracture

Measure Group	Measure Name (N=75)
Process	<b>OP-22</b> ED-Patient Left Without Being Seen
Process	<b>OP-23</b> ED-Head CT or MRI Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke who Received Head CT or MRI Scan Interpretation Within 45 Minutes of Arrival
Process	<b>PC-01</b> Elective Delivery Prior to 39 Completed Weeks Gestation: Percentage of Babies Electively Delivered Prior to 39 Completed Weeks Gestation
Process	<b>PN-6</b> Initial Antibiotic Selection for Community-Acquired Pneumonia (CAP) in Immunocompetent Patient
Process	<b>SCIP-Card-2</b> Surgery Patients on Beta-Blocker Therapy Prior to Arrival Who received a Beta-Blocker During the Perioperative Period
Process	<b>SCIP-Inf-1</b> Prophylactic Antibiotic Received Within One Hour Prior to Surgical Incision
Process	<b>SCIP-Inf-2</b> Prophylactic Antibiotic Selection for Surgical Patients
Process	<b>SCIP-Inf-3</b> Prophylactic Antibiotics Discontinued Within 24 Hours After Surgery End Time
Process	<b>SCIP-Inf-9</b> Urinary Catheter Removed on Postoperative Day 1 (POD 1) or Postoperative Day 2 (POD 2) with day of surgery being day zero
Process	<b>SCIP-VTE-2</b> Surgery Patients Who Received Appropriate Venous Thromboembolism Prophylaxis Within 24 Hours Prior to Surgery to 24 Hours After Surgery
Process	<b>STK-1</b> Venous Thromboembolism (VTE) Prophylaxis
Process	<b>STK-2</b> Discharged on Antithrombotic Therapy
Process	<b>STK-3</b> Anticoagulation Therapy for Atrial Fibrillation/Flutter
Process	<b>STK-4</b> Thrombolytic Therapy
Process	<b>STK-5</b> Antithrombotic Therapy By End of Hospital Day 2
Process	<b>STK-6</b> Discharged on Statin Medication
Process	<b>STK-8</b> Stroke Education
Process	<b>STK-10</b> Assessed for Rehabilitation
Process	<b>VTE-1</b> Venous Thromboembolism Prophylaxis
Process	<b>VTE-2</b> Intensive Care Unit Venous Thromboembolism Prophylaxis
Process	<b>VTE-3</b> Venous Thromboembolism Patients with Anticoagulation Overlap Therapy
Process	<b>VTE-4</b> Venous Thromboembolism Patients Receiving Unfractionated Heparin with Dosages/Platelet Count Monitoring by Protocol or Nomogram
Process	<b>VTE-5</b> Venous Thromboembolism Warfarin Therapy Discharge Instructions

Measure Group	Measure Name (N=75)
Process	<b><u>VTE-6</u></b> Hospital Acquired Potentially-Preventable Venous Thromboembolism
Efficiency	<b><u>OP-8</u></b> MRI Lumbar Spine for Low Back Pain
Efficiency	<b><u>OP-10</u></b> Abdomen Computed Tomography (CT) Use of Contrast Material
Efficiency	<b><u>OP-11</u></b> Thorax CT Use of Contrast Material
Efficiency	<b><u>OP-13</u></b> Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac Low-Risk Surgery
Efficiency	<b><u>OP-14</u></b> Simultaneous Use of Brain CT and Sinus CT

**Table 10. Measures excluded from overall star ratings (April 2015 data)**

Measure Name (N= 31)	Reason for Exclusion
<b><u>IMM-1a</u></b> Pneumococcal Immunization – Overall Rate	Measures suspended, retired, or delayed from public reporting
<b><u>SCIP-VTE-1</u></b> Surgery Patients with Recommended Venous Thromboembolism Prophylaxis Ordered	Measures suspended, retired, or delayed from public reporting
<b><u>AMI-2</u></b> Aspirin prescribed at discharge	Measures suspended, retired, or delayed from public reporting
<b><u>AMI-10</u></b> Statin prescribed at discharge	Measures suspended, retired, or delayed from public reporting
<b><u>HF-3</u></b> ACEI or ARB for LVSD	Measures suspended, retired, or delayed from public reporting
<b><u>SCIP-INF-4</u></b> Cardiac Surgery Patients With Controlled 6 A.M. Postoperative Blood Glucose	Measures suspended, retired, or delayed from public reporting
<b><u>SCIP-INF-10</u></b> Surgery patients with perioperative temperature management	Measures suspended, retired, or delayed from public reporting
<b><u>SM-PART-STROKE</u></b> Participation in a Systematic Clinical Database Registry for Stroke Care	Measures suspended, retired, or delayed from public reporting
<b><u>HF-1</u></b> Discharge Instructions	Measures suspended, retired, or delayed from public reporting
<b><u>PN-3b</u></b> Blood Cultures Performed in the ED Prior to Initial Antibiotic Received in Hospital	Measures suspended, retired, or delayed from public reporting
<b><u>OP-15</u></b> Use of Brain CT in the ED for Atraumatic Headache	Measures suspended, retired, or delayed from public reporting
<b><u>OP-29</u></b> Endoscopy/Polyp Surveillance: Appropriate Follow-Up Interval for Normal Colonoscopy in Average Risk Patients	Measures awaiting public reporting on <i>Hospital Compare</i>
<b><u>OP-30</u></b> Endoscopy/Polyp Surveillance: Colonoscopy Interval for Patients with a History of Adenomatous Polyps – Avoidance of Inappropriate Use	Measures awaiting public reporting on <i>Hospital Compare</i>
<b><u>OP-31 Cataracts</u></b> – Improvement in Patient’s Visual Function Within 90 Days Following Cataract Surgery	Measures awaiting public reporting on <i>Hospital Compare</i>
<b><u>CAC-3</u></b> Home Management Plan of Care (HMPC) Document Given to Patient/Caregiver	Measures with less than or equal to 100 hospitals reporting
<b><u>AMI-7a</u></b> Fibrinolytic Therapy Received Within 30 Minutes of Hospital Arrival	Measures with less than or equal to 100 hospitals reporting
<b><u>CAC-1</u></b> Relievers for Inpatient Asthma	Measures with less than or equal to 100 hospitals reporting
<b><u>CAC-2</u></b> Systemic Corticosteroids for Inpatient Asthma	Measures with less than or equal to 100 hospitals reporting

Measure Name (N= 31)	Reason for Exclusion
<b><u>OP-2</u></b> Fibrinolytic Therapy Received Within 30 Minutes of ED Arrival	Measures with less than or equal to 100 hospitals reporting
<b><u>OP-1</u></b> Median Time to Fibrinolysis	Measures with less than or equal to 100 hospitals reporting
<b><u>OP-9</u></b> Mammography Follow-up Rates	Non-directional efficiency measures
<b><u>MSPB-1/SPP-1</u></b> Medicare Spending per Beneficiary (MSPB)	Non-directional efficiency measures
<b><u>PAYM-30-AMI</u></b> Acute Myocardial Infarction (AMI) Payment per Episode of Care	Non-directional efficiency measures
<b><u>SM-PART-CARD</u></b> Participation in a Systematic Database for Cardiac Surgery	Structure measures without evidence of an association with changes in clinical practice or improved outcomes
<b><u>SM-PART-GEN-SURG</u></b> Participation in a Systematic Clinical Database Registry for General Surgery	Structure measures without evidence of an association with changes in clinical practice or improved outcomes
<b><u>SM-PART-NURSE</u></b> Participation in a Systematic Clinical Database Registry for Nursing Sensitive Care	Structure measures without evidence of an association with changes in clinical practice or improved outcomes
<b><u>ACS-REGISTRY</u></b> Participation in a multispecialty surgical registry	Structure measures without evidence of an association with changes in clinical practice or improved outcomes
<b><u>OP-12</u></b> The Ability for Providers with HIT to Receive Laboratory Data Electronically Directly into their ONC-Certified EHR System as Discrete Searchable Data	Structure measures without evidence of an association with changes in clinical practice or improved outcomes
<b><u>OP-17</u></b> Tracking Clinical Results between Visits	Structure measures without evidence of an association with changes in clinical practice or improved outcomes
<b><u>OP-25</u></b> Safe Surgery Checklist Use	Structure measures without evidence of an association with changes in clinical practice or improved outcomes
<b><u>OP-26</u></b> Hospital Outpatient Volume Data on Selected Outpatient Surgical Procedures	Structure measures without evidence of an association with changes in clinical practice or improved outcomes

## Appendix E. Hospital Summary Score Modeling Approach

We considered other options and their strengths and weakness as described below:

### ***Simple average***

While a simple average of all available measure scores for a hospital is easy to calculate and intuitive to understand, there are numerous limitations. This approach ignores the relationships between measures and assumes that all measures equally reflect underlying hospital quality.

### ***Weighted average of individual measures***

This approach is also generally straightforward to explain and intuitive by enabling certain measures to contribute more to the overall summary score. However, this approach also ignores the relationship between measures and requires an arbitrary, fixed assumption to be made about the relative contribution of each measure or measure type of quality.

### ***Latent variable models***

Latent variable models assume a latent “quality” trait at each hospital, which influences each hospital’s performance on the measures. The strength of the relationship between this latent quality trait, which is represented by the overall summary score, and each measure is called “loading”, and it is derived empirically from the data.

There are several advantages to the latent variable modeling approach.

- Explicitly models a single underlying quality trait that might influence quality measures; such a latent trait is an implicit assumption of a single dimensional star rating.
- There is a statistical evidence base for the application of this modeling approach for aggregating healthcare quality measures<sup>1</sup> as well as in other disciplines such as psychology and education.
- Accounts for the relationship, or correlation, between measures within hospitals by using a random effect (latent variable).
- The value of latent variable for each hospital can be estimated and is the summary score of that hospital. This estimate has the statistical property of being the Best Linear Unbiased Prediction (BLUP)<sup>2</sup> for the latent variable.
- Loadings are estimated from the model which is derived empirically (i.e. based on computations using observed data and not subjectively assigned). Therefore, how much a measure contributes to the overall summary score is determined by how consistently it varies with all of the other

---

<sup>1</sup> Landrum, Mary Beth, Bronskill, Susan E, Normand, Sharon-Lise. Analytic Methods for Constructing Cross-Sectional Profiles of Health Care Providers, Health Services & Outcomes Research Methodology 1:1 (2000): 23±47

<sup>2</sup> Henderson, C.R. (1975) “Best linear unbiased estimation and prediction under a selection model” Biometrics 31(2):423-447

measures. “Consistent” measures give a stronger signal about a hospital’s quality profile than “random” performance.

- The loadings for individual measures dynamically change as the distribution of hospital performance on individual measures and the relationship between individuals measures evolve over reporting periods.

However, several limitations and assumptions of this approach are also worthy of mention prior to detailed discussion.

- Empirically calculated loadings may not match expected or traditional conceptual frameworks of measure importance or contribution to quality.
- The ordinal ranking of the relationship between individual measures and summary score will vary over time.
- As no gold standard for overall hospital quality exists upon which to validate or “check the model,” we need to assess the face validity of numerous assumptions at each stage of model development. Notably, this limitation is common to all of the described approaches.

## Appendix F. Two-Stage Approach with Measure-Type Groups

In response to feedback from the TEP, the public comments, and input from partners within CMS, CORE proposes a two-stage approach to calculate hospital summary scores in which measures are categorized into groups for initial modeling, and group scores are subsequently aggregated using purposeful (conceptual) weights. CMS has reviewed this approach and believes it will meet policy objectives and effectively incorporate stakeholder feedback, while fulfilling the ultimate project goal to make *Hospital Compare* more accessible and useable to patients and consumers. CMS would also like any additional TEP input on this approach and its modifications.

### ***Proposed Grouping Method***

CORE evaluated several conceptual options or dimensions of quality and proposes organizing quality measures into six mutually exclusive conceptual groups by measure type for the following reasons.

- The six measure-type groupings are analogous to the CMS HVB program and other national quality initiatives.
- This decision was generally supported during the previous TEP meeting.
- Each group accounts for measures that represent similar concepts of quality.
- Measure-type groups are clinically reasonable in that they capture common components of quality for which hospital quality is likely linked across measures. For example, the degree to which hospitals effectively provide care or safely discharge patients is likely to be reflected in all readmission measures. Note: this assumption is likely less strong for the Process measure group.
- Each measure group can be reliably replicated as measures are added and removed from the star ratings. Other options for grouping may not permit the mutually exclusive assignment of measures to each group or be more subjective and would require substantially more initial and subsequent deliberation and compromise to ensure consistency.
- Utilized by many other CMS and non-CMS programs to identify separate conceptual dimensions of quality.

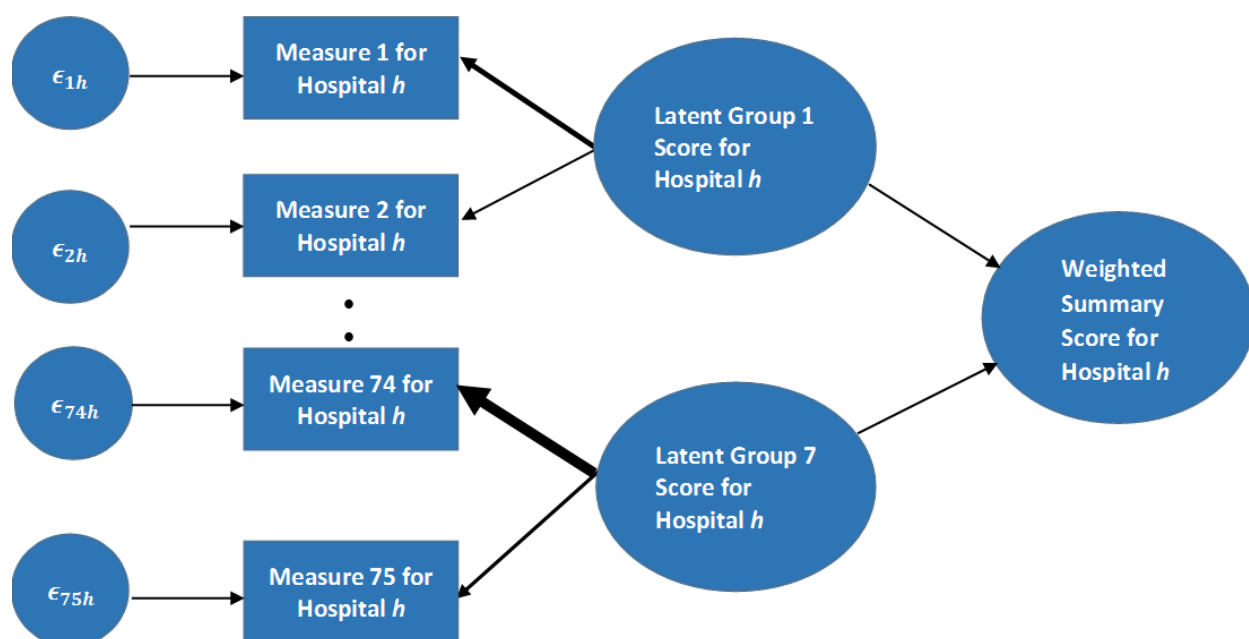
**Figure 2. Proposed Measure-Type Groups (December 2015 data)**



## ***Proposed Approach: Two-Stage using Latent Variable Model and Weighted Averaging***

The Two-Stage approach with measure groups combines the summary scores of six measure-type groups into one single overall hospital summary score. In step one, six separate latent variable models are generated for each measure-type group to calculate a group-specific latent summary score. The group-specific latent summary scores are standardized. In step two, a policy-based weighting scheme is applied to calculate a weighted average of the group-specific latent summary scores to generate the overall hospital summary score.

**Figure 3. Path Diagram for Example Two-Stage Approach with Measure Groups**



### **Advantages of Using Measure Groups**

- Groups account for measures that represent similar concepts of quality
- Use of groups aligns Overall Hospital Quality Star Ratings with other CMS star efforts
- Use of weights in second stage can flexibly incorporate policy priorities and preferences
- Reduces need to consider grouping individual or select measures

### **Limitations of Using Measure Groups**

- Overall summary score requires a subjective, policy-based weighting scheme
- Relationship between individual measures in different groups is not reflected in loadings

## ***Use of LVM in Two-Stage Approach***

The first stage of the Two-Stage Approach applies LVM to each of the six measure-type groups. LVM seeks to measure the unobserved, group-specific aspect of quality that is reflected by a hospital's performance on measures within a group. In addition to LVM, CORE considered other options for calculating group-specific summary scores (e.g., weighted average).

### **Advantages of LVM to Calculate Group-Specific Summary Scores**

- Method is used for composite measures in health quality literature.<sup>3</sup>
- Method is reproducible and feasible in comparison to weighted average of individual measures to generate a group score.
- Loadings account for consistency (or correlation) of measures within a group.
- LVM accounts for missing measures by using only available information to generate a group score. Hospitals with few measures are pulled towards the mean.
- Sampling variance can be accounted for within the model.
- Confidence intervals can be calculated around the overall summary scores.

### **Limitations of LVM to Calculate Group-Specific Summary Scores**

- Model complexity may be challenging for patients and consumers to understand.
- Other star ratings efforts use different methodologies for combining measures potentially leading to differences in rankings.
- Use of latent variable model for patient experience domain could result in similar, but different patient experience summary scores than HCAHPS Star Ratings.

---

<sup>3</sup> Schwartz M, Ren J, Pekoz EA, Wang X, Cohen AB, Restuccia JD. Estimating a composite measure of hospital quality from the Hospital Compare database: differences when using a Bayesian hierarchical latent variable model versus denominator-based weights. *Medical care* 2008;46:778-85.

## Appendix G. Proposed Weighting Options for Two-Stage Approach

### ***Purposeful Weighting***

The second stage of the Two-Stage Approach applies weights to summarize six group-level scores into one single overall summary score. It requires the development of a weighting scheme to generate hospital summary scores. Ideally, this weighting scheme would represent the preferences of stakeholders, particularly patients and consumers.

CORE has developed two potential options for presentation to CMS and the TEP:

- Equal weighting
- Weighting modified from the FY 2017 Hospital Value Based Purchasing (HVPB) program

In addition to this TEP meeting, these weighting options will also be vetted with CMS partners, the public via a second public comment period and through our partnership with the National Partnership for Women and Families.

***The primary purpose of TEP discussion is to review the relative and proportional importance of each measure domain for weighting, not to select a specific numerical weight.***

Prior to the TEP meeting on March 30, CORE will distribute a brief survey regarding the weighting of these groups that will be used to facilitate the TEP conversation.

**Table 11. Purposeful Weighting Scheme Modified from Hospital Value-Based Purchasing**

Conceptual Group	FY17 HVPB Weight	Proposed Weight
Outcomes – Safety (N=8)	20%	22%
Outcomes – Readmission (N=7)	--	22%
Outcomes – Mortality (N=6)	25%	22%
Patient Experience (N=11)	25%	22%
Process – Clinical Care (N=30)	5%†	4%
Process – Timeliness (N=8)	--	4%
Efficiency – Resource Use (N=5)	--	4%
Efficiency – Cost	25%	--

Note: The Hospital FY17 VBP program is set to include a total of 22 measures and assigns a 25% weight to the Medicare Spending per Beneficiary measure, which is not included in Star Ratings.

## Appendix H. Proposed Options for Star Ratings Translation

The decision of how to translate hospital summary scores into stars is largely policy-based. After consultation with CMS, there are several important assumptions to consider prior to evaluating each option.

- There will always be hospitals at the higher and lower ends of the distribution; some hospitals will fall into the one-star category and some will fall into the five-star category.
- Similar to other CMS Star Ratings efforts, a three-star rating will be “average.”
- The objective of this project is to develop whole-star ratings (not half-stars).
- Star ratings do not reflect an “apples to apples” comparison between hospitals, but rather reflect the weighted average of summarized, group-level quality information available for a given hospital.
- Star ratings are not intended to guide hospital quality improvement efforts, but rather to make summary information available to the public.

Three options we presented for TEP feedback include:

- Fixed Intervals;
- Overall and Measure-specific Threshold; and
- Clustering.

### ***Option 1: Fixed Interval***

The fixed interval method for translating summary scores to stars is determined by a hospital’s percentile rank (i.e., the rank of a hospital’s summary score).

**Table 12. Example of Fixed Interval Star Categories**

Stars	Rating
Five stars	Hospital’s summary score is larger than the 80 <sup>th</sup> percentile and lower than or equal to the 100 <sup>th</sup> percentile.
Four stars	Hospital’s summary score is larger than the 60 <sup>th</sup> percentile and lower than or equal to the 80 <sup>th</sup> percentile.
Three stars	Hospital’s summary score is larger than the 40 <sup>th</sup> percentile and lower than or equal to the 60 <sup>th</sup> percentile.
Two stars	Hospital’s summary score is larger than the 20 <sup>th</sup> percentile and lower than or equal to the 40 <sup>th</sup> percentile.
One star	Hospital’s summary score is lower than or equal to the 20 <sup>th</sup> percentile.

### Advantages of Fixed Interval Approach

- Categories can be set to ensure a specific proportion of hospitals in each category.
- Simplicity of the approach may be easy for patients and consumers to understand.

### Limitations of Fixed Interval Approach

- Categories may not be statistically different;
  - For example, no meaningful difference between 20<sup>th</sup> and 21<sup>st</sup> percentiles.

### ***Option 2: Overall and Measure-Specific Threshold***

Setting a threshold for a hospital's performance on both the summary score and individual measure performance is similar to an approach employed by other CMS efforts. This approach is based on categorizing each star rating based on the statistical difference in both summary score and individual measure performance.

In this approach, the threshold for individual measure performance can be set based on face validity and policy objectives.

**Table 13. Example of Overall and Measure-Specific Threshold Star Categories**

Stars	Rating
Five stars	Hospital's summary score is significantly better than the national average and more than half of the individual measures are better than the national average.
Four stars	Hospital's summary score is significantly better than the national average and less than half of the individual measure are better than the national average.
Three stars	Hospital's Summary score is not significantly different form the national average or no other criteria is met.
Two stars	Hospital's summary score is significantly worse than the national average and less than half of the individual measure are worse than the national average.
One star	Hospital's summary score is significantly worse than the national average and more than half of the individual measure are worse than the national average.

### Advantages of Overall and Measure Specific Threshold Approach

- Categories reflect statistical differences in the summary score
- Hospitals with few measures must perform well on the same percentage of total measures reported as hospitals with many measures

### Limitations of Overall and Measure Specific Threshold Approach

- Frequency of hospitals in each star category are likely to be unequal
- Majority of hospitals falling into the three-star category
- Requires face validity to set threshold
- May appear to “double count” individual measure performance

### Option 3: *k*-Means Clustering

*k*-means cluster analysis is a method for creating groups (or clusters) so that observations in each cluster are closer to their group mean than to any other group mean. In the case of a summary score, *k*-means cluster analysis with  $k=5$  can be used to categorize hospitals into a star rating category such that hospitals in each star category have an overall score that is ‘more like’ the other hospitals in that star category than it is like hospitals in different categories.

This approach is based solely on the hospital summary score and does not re-assess individual measure performance.

**Table 14. Example of *k*-Means Clustering Star Categories**

Stars	Rating
Five stars	Cluster of hospitals with the highest star determined by minimizing the sum of the square of distance between hospital’s summary scores within the cluster
Four stars	Cluster of hospitals with the second highest star determined by minimizing the sum of the square of distance between hospital’s summary scores within the cluster
Three stars	Cluster of hospitals with average star determined by minimizing the sum of the square of distance between hospital’s summary scores within the cluster
Two stars	Cluster of hospitals with the second lowest star determined by minimizing the sum of the square of distance between hospital’s summary scores within the cluster
One star	Cluster of hospitals with the lowest star determined by minimizing the sum of the square of distance between hospital’s summary scores within the cluster

### **Advantages of *k*-Means Cluster Approach**

- Hospitals in a cluster have statistically similar summary scores
- Individual measure performance does not appear to be assessed “twice” as in the threshold approach

### **Limitations of *k*-Means Cluster Approach**

- Majority of hospitals fall into three-star cluster
- Complexity of approach may be difficult for patients and consumers to understand

## Appendix I. Approach for Missing Measures or Measure Groups

Two important considerations that will impact the assignment of hospital summary scores relate to incorporating hospitals with missing measures and/or missing domains.

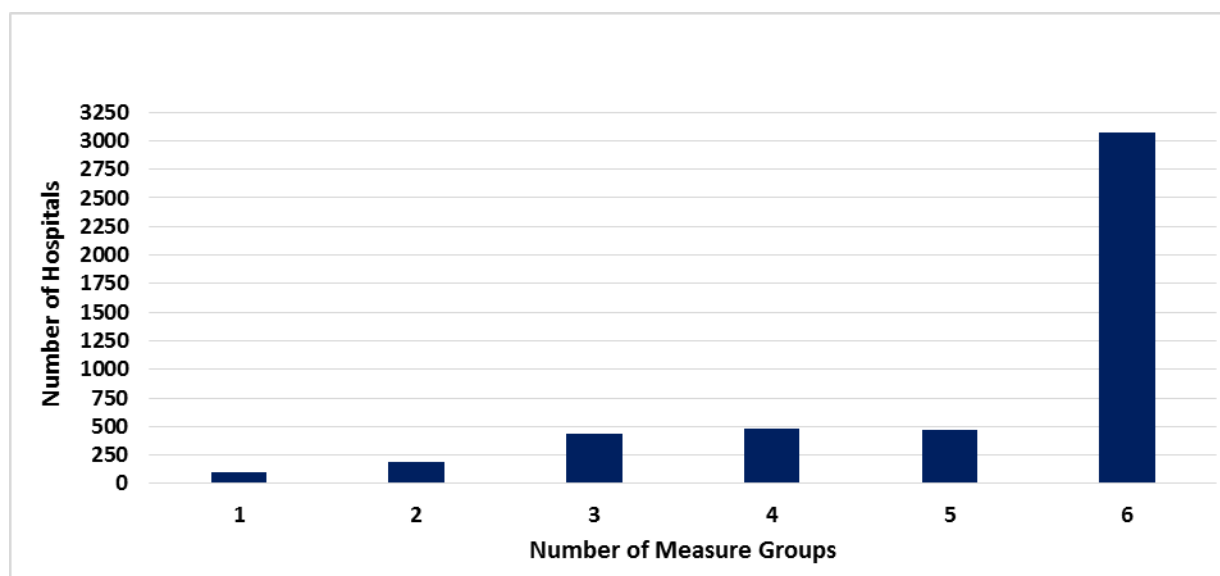
### ***Minimum Number of Measures***

A hospital must publicly report a minimum number of measures for it to receive an overall hospital star rating. Given the variable number of measures each hospital reports within each measure group, an approach to setting a minimum threshold that carries face validity must be developed.

### ***Minimum Number of Domains***

Use of the two-stage approach using LVM and weighted averaging will also generate scenarios where hospitals have different numbers of measure groups with available hospital quality information ([Figure 4](#). Distribution of Number of Measure Groups per Hospital). As the Overall Hospital Star Ratings seeks to only reflect available information about each hospital, and not impute a hypothetical value, both measure group thresholds and methods to reapportion conceptual weights must be developed.

**Figure 4. Distribution of Number of Measure Groups per Hospital**



### ***Recommendation for Minimum Thresholds for Public Reporting***

After consultation with CMS, CORE seeks TEP input regarding minimum measure and measure group thresholds for the public reporting of star ratings.

As a point of reference, the HVB program currently requires a minimum of 4 measures for the Process measure group, 100 surveys for HCAHPS measures (same as star ratings), 2 measures for the Outcome measure group, and at least 2 of 4 domains to have a Total Performance Score calculation.

One primary tradeoff associated with setting increasingly high thresholds for both measures and measure-type groups will be excluding more hospitals from the Overall Hospital Quality Star Ratings.

[Table 15](#) reflects changes in the number of hospitals with a reported star rating based on individual measure and domain thresholds.

**Table 15. Hospitals with Summary Score by Minimum Measure and Measure Group Thresholds (December 2014 data)**

Minimum Measures/Group	Minimum Groups			
	3	4	5	6
2	4,135 (87%)	3,697 (78%)	3,272 (69%)	2,789 (59%)
3	3,619 (76%)	3,095 (65%)	2,675 (56%)	2,246 (47%)
4	2,886 (61%)	2,504 (53%)	2,091 (44%)	1,898 (40%)

Note: The minimum group threshold also assumes a minimum of one outcome group available for the star ratings. The total number of hospitals eligible for an overall star rating is 4,753 hospitals. Please note that 292 hospitals in the December 2014 *Hospital Compare* dataset report no measures and therefore are not included in these results.

### ***Recommendation for Weighting***

With respect to the weighting of measure groups when groups are missing, CMS seeks further input from the TEP. As an example, CORE has proposed following the approach taken by the HVB program in which weights of missing domains are re-proportioned across the domains that a hospital reports.

An example adjusted weighting scheme accounting for missing measure groups is shown in [Table 16](#).

**Table 16. Example of Re-weighting Scheme when Missing Outcome – Safety Measures**

Measure-Type Group	Proposed Weight Reviewed by TEP	Re-proportioned Weight
Patient Experience (N=11)	25	31.25
Outcome – Mortality (N=4)	25	31.25
Outcome – Readmissions (N=5)	20	25
Outcome – Safety (N=0)	20	--
Process (N=38)	5	6.25
Efficiency (N=5)	5	6.25

## Appendix J. TEP Minutes

### Hospital Quality Star Ratings on Hospital Compare Technical Expert Panel

#### Summary Call #1

Monday, December 1, 2014, 5:00-7:00 PM ET

#### PARTICIPANTS.

**Technical Expert Panel (TEP):** Matt Austin, PhD; Vinita Bahl, DMD, MPP; John Bott, MBA, MS; Kathy Ciccone, RN, MBA; Kelly Court, MBA; Rachel Grob, PhD; Emma Kopleff, MPH; Doris Peter, PhD; Casey Schwarz, JD; David Shahian, MD; Brett Stauffer, MD, MHS; Guofen Yan, PhD; Ben Yandell, PhD

**Yale New Haven Health Services Corporation – Center for Outcomes Research (CORE):** Arjun Venkatesh, Jaymie Potteiger, Susannah Bernheim, Mallory Perez, Ben Clopper, Zhenqui Lin, Haiqun Lin, Angela Hsieh, and Nirupama Krishnamurthi

**Lantana Consulting Group (Lantana):** Kit Cooper, Hector Cariello

**National Opinion Research Center at the University of Chicago (NORC):** Romonda Bumpus

**Centers for Disease Control and Prevention (CDC):** Jonathan Edwards, Dawn Sievert

**Centers for Medicare & Medicaid Services (CMS):** Lein Han, Kristie Baus, Bill Lehrman, Pierre Yong, Elizabeth Goldstein

SUMMARY	ACTION ITEMS
<ul style="list-style-type: none"> <li>• Welcoming remarks and introductions</li> <li>• TEP goals and introductions <ul style="list-style-type: none"> <li>○ Capture constituent interests and gain input from a broad set of stakeholders through several avenues.</li> </ul> </li> <li>• TEP charter and timeline <ul style="list-style-type: none"> <li>○ Next TEP meetings to be held in January and February with public comment periods most likely being held in January and March.</li> <li>○ Proposed dry run in Summer 2015</li> </ul> </li> <li>• Project status <ul style="list-style-type: none"> <li>○ Criteria review</li> <li>○ Measure inclusion</li> </ul> </li> <li>• Criteria: measures recommended for exclusion</li> <li>• Criteria: measures requiring detailed TEP evaluation</li> <li>• Next steps</li> </ul>	<ul style="list-style-type: none"> <li>• <b>TEP members should reach out to the CORE team with any major concerns with any of the exclusions or inclusion criteria discussed during the meeting.</b></li> <li>• <b>The majority of the TEP supported the three exclusion criteria proposed by CORE :</b> <ul style="list-style-type: none"> <li>○ <b>Measures not publicly reported on Hospital Compare due to suspension of data collection</b></li> <li>○ <b>Measures awaiting public reporting of performance data on Hospital Compare in the future</b></li> <li>○ <b>Measures with less than or equal to 100 hospitals reporting</b></li> </ul> </li> </ul>

SUMMARY	ACTION ITEMS
	<ul style="list-style-type: none"> <li>Discussion regarding retaining measures that are de-endorsed by NQF and not supported by the MAP will be continued. <ul style="list-style-type: none"> <li>The star ratings will seek to be consistent with what remains on <i>Hospital Compare</i>.</li> </ul> </li> <li>CORE will bring TEP feedback to CMS and consider several factors regarding including/excluding “topped out” measures in the summary star. CORE will update the TEP with CMS’s decision. <ul style="list-style-type: none"> <li>The star ratings approach will seek to be consistent with what remains on <i>Hospital Compare</i>.</li> </ul> </li> <li>The TEP generally supported the exclusion of the registry measures, the current structural measures of volume, and other dichotomous structural measures (OP-12 The Ability for Providers with HIT to Receive Laboratory Data Electronically Directly into their ONC-Certified EHR System as Discrete Searchable Data, OP-17 Tracking Clinical Results between Visits, and OP-25 Safe Surgery Checklist) without evidence of an association with changes in clinical practice or improved outcomes.</li> <li>CORE will bring the subject of using select procedural volumes from OP-26 Hospital Outpatient Volume Data on Selected Outpatient Surgical Procedures back to the TEP for additional input if technically feasible.</li> </ul>

TOPIC	GENERAL
<b>Welcoming Remarks and Introductions</b>	<p>Jaymie Potteiger, Project Coordinator, MPH, CORE, gave a brief welcome and introduction on behalf of the Centers for Medicare and Medicaid Services (CMS) and the Center for Outcomes Research and Evaluation (CORE).</p> <p>Ms. Potteiger asked that everyone at the meeting keep all personal opinions and experiences shared confidential, and all project decisions and results will remain confidential until publicly reported by CMS. She reviewed the agenda, which included introductions, project overview, measure selection, and next steps.</p> <p>Ms. Potteiger introduced the CMS representatives on the task. Kristie Baus, RN, CMS, is the contracting officer representative for the Lantana work. Lein Han, PhD, CMS, is the</p>

TOPIC	GENERAL
	<p>contracting officer representative for the CORE work. Pierre Yong, MD, MPH, CMS, is the director of the Division of Hospitals and Medication Measurement.</p> <p>Ms. Potteiger introduced the CMS contractors working on the Star Ratings task and mentioned that CORE and Lantana are developing the methodology for the star ratings that summarize the quality measures currently on <i>Hospital Compare</i>.</p> <p>Ms. Potteiger provided an overview and introduced the CORE team.</p> <p>Kit Cooper, Project Manager, Lantana, provided an overview for Lantana and introduced the Lantana and NORC teams.</p>
TEP Goals and Introductions	<p>Arjun Venkatesh, MD, MBA, MHS, Project Lead, CORE, thanked the TEP members for participating in the TEP process and explained that the goal was to capture constituent interests and gain input from a broad set of stakeholders through several avenues.</p> <p>Dr. Venkatesh asked the TEP members to introduce themselves to the group and disclose any conflicts of interest, including both conflicts previously disclosed in the application process or any new conflicts.</p> <p>Matt Austin, PhD, Assistant Professor at the Armstrong Institute for Patient Safety and Quality, Johns Hopkins University mentioned he has a contract with the Leapfrog Group and receives funding through his Commonwealth Fund.</p> <p>Vinita Bahl, DMD, MPP, Director of Performance Assessment and Clinical Effectiveness, University of Michigan Health System reported no conflicts of interest to disclose.</p> <p>John Bott, MBA, MS, Manager of Performance Measurement, State of Wisconsin Department of Employee Trust Funds and consultant with <i>Consumer Reports</i> reported no conflicts of interest to disclose.</p> <p>Kathy Ciccone, RN, MBA, Executive Director for the Quality Institute for the Healthcare Association of New York State reported no conflicts of interest to disclose.</p> <p>Kelly Court, MBA, Chief Quality Officer at the Wisconsin Hospital Association reported no conflicts of interest to disclose.</p> <p>Rachel Grob, PhD, Director of National Initiatives at the Center for Patient Partnerships at the University of Wisconsin-Madison reported no conflicts of interest to disclose.</p> <p>Dr. Venkatesh mentioned that Rodney Hayward, MD was not available to join the call.</p> <p>Emma Kopleff, MPH, Senior Policy Advisor, National Partnership for Women &amp; Families reported no conflicts of interest to disclose.</p> <p>Doris Peter, PhD, Director, <i>Consumer Reports</i> Health Ratings Center reported that she uses CMS quality data to make composites as a potential conflict of interest.</p>

TOPIC	GENERAL
	<p>Dr. Venkatesh mentioned that Lauren Petersen, MD, MPH was not available to join the call.</p> <p>Casey Schwarz, JD, Client Services and Policy Counsel at the Medicare Rights Center reported no conflicts of interest to disclose.</p> <p>David Shahian, MD, Vice President for Quality and Safety at Massachusetts General Hospital and Professor of Surgery at Harvard Medical School stated that he also chairs the Society of Thoracic Surgeons' national database and serves on the board of directors of the National Quality Forum (NQF) and executive committee of the <b><i>Physician Consortium for Performance Improvement</i></b> (PCPI). Dr. Shahian said that all of these positions, with the exception of Massachusetts General Hospital, are voluntary and unpaid. He reported no conflicts of interest to disclose.</p> <p>Brett Stauffer, MD, MHS stated he is moving into the role of Vice President of Hospital Care Quality Improvement for Baylor Scott and White Health. He reported no conflicts of interest to disclose.</p> <p>Guofen Yan, PhD, Associate Professor, School of Medicine at the University of Virginia reported no conflicts of interest to disclose.</p> <p>Ben Yandell, PhD, Associate Vice President, Norton Healthcare reported no conflicts of interest to disclose.</p> <p>Dr. Venkatesh stated the goal of the project is to improve the usability and interpretability of <i>Hospital Compare</i> for consumers and patients by developing a scientifically valid summary star rating methodology. He added that this work is part of a larger CMS effort to improve the transparency of hospital quality information as well as consumer and patient engagement. Dr. Venkatesh acknowledged there are many inherent limitations to developing a five-star system, but that it should not preclude the team from trying to make the information more available to the public.</p> <p>Dr. Venkatesh reviewed the role of the TEP as providing expert feedback to CMS to inform the development Star Ratings. In addition, Dr. Venkatesh notes that the TEP is one of several avenues being utilized by CMS to gather stakeholder input and to ensure transparency in the development process. Dr. Venkatesh framed each TEP meeting as an opportunity for TEP members to think about how we can address some of the major methodological issues and decisions with creative and innovative ideas in order to create a star rating that is both scientifically sound but also accessible to patients.</p>

TOPIC	GENERAL
<b>TEP Charter and Timeline</b>	<p>Ms. Potteiger reviewed the TEP charter and outlined the main goals. She explained that appointment to the TEP for this project runs through September 2015 and asked if anyone had any concerns or comments about the charter. There were no comments or concerns voiced, and thus, the charter was considered approved.</p> <p>Ms. Potteiger reviewed the goals for each TEP meeting and the project timeline and mentioned the next TEP meetings would be held in January and February with public comment periods most likely occurring in January and March. She mentioned that there is a proposed dry run in Summer 2015. CMS is considering implementing the star ratings on <i>Hospital Compare</i>, but no final decisions have been made.</p>

TOPIC	DISCUSSION	ACTION ITEM
<b>Project Status</b>	<p>Dr. Venkatesh reviewed the work previously completed by the development teams including the Environmental Scan and Literature Review (ES/LR) and other collaboration and coordination efforts.</p> <p>Dr. Venkatesh explained that measure selection has been broken into two levels for this discussion, criteria review and measure inclusion review, and the main focus of the discussion should be on criteria review. He provided an overview of the process used to refine the measures for which the team needs the TEP's input most. There were 33 measures for discussion to include or exclude from the summary star rating.</p> <p>Dr. Venkatesh reviewed the three criteria for exclusion. He explained that measures that have been retired or delayed from public reporting do not have updated, publicly available data on <i>Hospital Compare</i>. Similarly, measures awaiting future public reporting cannot be included in the star ratings until publicly available data is accessible through <i>Hospital Compare</i>. The teams proposed that the summary star rating should only include data that has already been open to the public. He provided an explanation for reasons to include or exclude measures where <math>\leq 100</math> hospitals have a performance score. Dr. Venkatesh then sought feedback from the TEP about the recommended exclusion measures.</p> <p>One TEP member asked if the construct for the star ratings had been created and if quality had been defined. Dr. Venkatesh stated that it has been difficult to draw boundaries around what defines quality. The</p>	<p><b>TEP members should reach out to the CORE team with any major concerns with the three exclusion criteria.</b></p> <p><b>Discussion of retaining measures that are de-endorsed and not supported by MAP will be continued.</b></p> <p><b>CORE will bring the subject of using select procedural volumes from OP-26 back to the TEP for additional input.</b></p> <p><b>Exclude OP-12.</b></p>

TOPIC	DISCUSSION	ACTION ITEM
	<p>team seeks to summarize a hospital's performance across currently available measures into a single score.</p> <p>Another TEP member asked for an explanation of why retired measures are proposed for exclusion from the summary star rating if the purpose of the star ratings is to provide the consumer with a sense of the overall quality of the hospital. Dr. Venkatesh stated that retired measures do not have updated data available on <i>Hospital Compare</i>, which presents a challenge for inclusion. He added that some of these measures are retired because evidence indicated that they were no longer scientifically valid. He agreed that the potential reasons for measure retirement can be included in the materials for public comment.</p> <p>One TEP member asked how the star ratings would be updated if an individual measure meeting all of the inclusion/exclusion criteria is introduced or retired from public reporting. Additionally, the TEP member asked how the team plans to communicate this update to end users of <i>Hospital Compare</i>. Dr. Venkatesh indicated that the team has not yet set a definite plan for re-evaluation of the star ratings, but will ensure this topic is included in the discussion of a future TEP meeting.</p> <p>One TEP member agreed with all three of the recommended exclusion criteria as long as, for the measures retired or delayed from public reporting, the star rating aligns with the measures currently on <i>Hospital Compare</i>. For measures with less than or equal to 100 hospitals with a performance score, the TEP member suggested setting a minimum number of measures required for each hospital in order to create a valid rating. Dr. Venkatesh stated that once the measures have been determined, the team will bring back data to the second TEP for further discussion on the minimum number of measures required of hospitals for reporting a summary star rating. The TEP member asked if the star rating would include only measures currently reported on <i>Hospital Compare</i>. Dr. Venkatesh confirmed that all measures included in the summary star rating will be reported on <i>Hospital Compare</i>; however, <i>Hospital Compare</i> may also report measures separate from the star ratings.</p> <p>One TEP member stated that the exclusion criteria may limit the information provided on aspects of quality that patients value. The TEP member asked if this TEP could provide recommendations about aspects of quality not captured by existing measures that they would</p>	<p><b>Contact Dr. Venkatesh if you disagree that OP-17 should be excluded.</b></p> <p><b>Strategies for including/excluding topped-out measures will be addressed in future TEP meetings or methodology reports.</b></p>

TOPIC	DISCUSSION	ACTION ITEM
	<p>like to be represented in the star ratings if new measures were developed. Dr. Venkatesh agreed and welcomed individual feedback about major measurement gaps.</p> <p>Dr. Venkatesh encouraged TEP members to reach out to the CORE team with any major concerns with the three exclusion criteria and moved the discussion to the four exclusion criteria requiring TEP input: measures that have been de-endorsed by the NQF and are no longer supported by the Measure Applications Partnership (MAP), structural measures, topped-out measures, and efficiency measures.</p> <p>Dr. Venkatesh reviewed the survey results for the measure de-endorsed by NQF and not supported by the MAP and outlined the reasons for exclusion and inclusion of these measures. He opened the discussion to the group and asked for input on whether the decision for inclusion or exclusion can be made at the criteria level or if a measure-by-measure discussion was needed.</p> <p>Another TEP member asked for clarification on the target population for the star ratings, all potential consumers and patients using <i>Hospital Compare</i> or only Medicare beneficiaries. Dr. Venkatesh suggested that the TEP consider the broader view as much as possible.</p> <p>One TEP member asked if measures that are de-endorsed by the NQF and not supported by the MAP are likely to be dropped from the <i>Hospital Compare</i> website. Dr. Venkatesh stated that retirement from <i>Hospital Compare</i> is likely but not necessarily a rule or requirement. Kristie Baus stated that CMS looks at a variety of criteria to determine if a measure should remain on the website and there is a tendency, but not a guarantee, to follow what the MAP recommends.</p> <p>One TEP member supported measures de-endorsed by NQF and not supported by the MAP as an exclusion, but clarified that “not supported by the MAP” is meant to refer to the MAP’s guidance for public reporting.</p> <p>Another TEP member stated that a measure-by-measure discussion creates a difficult precedent because measures continue to change and the star ratings should align with what CMS approves for public reporting.</p> <p>Susannah Bernheim MHS, MD, CORE, restated a TEP member’s comments for clarification: if a measure has been de-endorsed by NQF</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>and does not have the MAP's support, but CMS finds it valuable enough to display on <i>Hospital Compare</i>, the measure should be included in the star rating so there is concordance between what is publicly reported and what is incorporated into the summary star.</p> <p>One TEP member stated that from the consumer viewpoint, there is some concern that if CMS is retaining measures that are de-endorsed and not supported by MAP, the consumer voice is underrepresented and these measures would not be truly representative.</p> <p>Dr. Venkatesh stated that the discussion was helpful and concluded that the inclusion or exclusion of these will be evaluated by CMS based on this TEP input as well as the need to maintain consistency with existing reporting programs.</p> <p>Dr. Venkatesh reviewed the survey results for exclusion and inclusion for structural measures with dichotomous outcomes and outlined reasons for exclusion and inclusion of these measures. He opened the discussion to the group and asked whether, at the criteria level, a registry measure with an answer of yes or no participation should be included in a hospital's summary star rating.</p> <p>One TEP member asked for clarification if the TEP should offer input on the six to seven dichotomous structural variables specifically listed or consider the criteria as a whole. Dr. Venkatesh stated that the team would first like feedback on measures of registry participation. He further clarified the question to ask if the TEP thinks registry participation measures with yes/no responses should be part of the summary star rating. He added that the discussion today should consider registry measures as they are currently specified and reported.</p> <p>Another TEP member stated that they would exclude these measures since participation in a registry is simply data transmission and not a marker of quality and hospitals' responses are not validated.</p> <p>Dr. Venkatesh reviewed the survey results for structural measures of procedural volume and outlined reasons for exclusion and inclusion of these measures. He opened the discussion to the group.</p> <p>One TEP member stated that they felt having volume measures available is important, but said they do not agree with including them as part of the star ratings. The TEP member added that inclusion might</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>result in an unintended consequence of doing more procedures that potentially do not need to be done.</p> <p>Another TEP member agreed that procedural volume measures should be excluded from star ratings.</p> <p>One TEP member added that for the procedural volume measures that have been well-vetted, the evidence shows a direct association between outcomes and procedural volume. Dr. Venkatesh stated that the data available reflects each hospital's Medicare fee-for-service beneficiary volume only. The volume measures under consideration for the star ratings is constrained to the set of procedures captured in OP-26 Hospital Outpatient Volume Data on Selected Outpatient Surgical Procedures as it is the only volume measure currently reported on <i>Hospital Compare</i>. Dr. Venkatesh stated that for the purposes of this TEP, the TEP is limited to what has already been specified and reported.</p> <p>Dr. Bernheim summarized the TEP's recommendation as procedural volume measures should be included in star ratings if substantial evidence of an association between volume and outcomes exists.</p> <p>One TEP member agreed with Dr. Bernheim's summary, adding that the team would also have to consider a hospital's case mix before including a procedural volume measure in the summary star rating. If the procedure volume measure is specific to the Medicare population, the TEP member suggests there be a minimum volume requirement for Medicare patients.</p> <p>Dr. Venkatesh stated that not all procedures included in OP-26 have evidence of a volume-outcome association; however, certain procedure subsets of this measure where the volume-outcome relationship is supported may be meaningful for the star ratings. CORE will bring a list of these procedures/procedure subsets back to the TEP for additional input.</p> <p>Another TEP member voiced concern that this information is useful for hospitals and patients but supported exclusion from the star summary.</p> <p>Dr. Venkatesh reviewed the three other structural measures with dichotomous outcomes -- OP-12 The Ability for Providers with HIT to Receive Laboratory Data Electronically Directly into their ONC-Certified EHR System as Discrete Searchable Data, OP-17 Tracking Clinical Results between Visits, and OP-25 Safe Surgery Checklist -- and discussed the</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>survey results. He asked the TEP if they had any strong feelings for the inclusion of OP-12 and received no response. Dr. Venkatesh noted this as approval to exclude this measure.</p> <p>Dr. Venkatesh asked the TEP if there was strong feeling for or against including OP-17 in the summary star. One TEP member stated they did not feel this was a good measure to include, adding that it will be difficult to determine the hospital's use of the electronic health record (EHR). Dr. Venkatesh asked the TEP if anyone felt strongly that this measure conveys important information to patients and consumers and therefore should be included. There was no response from the TEP members.</p> <p>Dr. Venkatesh asked the TEP if there was strong feeling for or against including OP-25 in the summary star. There was discussion that the use of the checklist does not indicate information about a hospital's quality. Three of the TEP members agreed it should not be included in the summary star. Dr. Venkatesh asked anyone who felt strongly that this measure should be included to contact him; otherwise, the general sentiment was not to include this measure.</p> <p>Dr. Venkatesh reviewed the survey results for "topped out" measures and outlined reasons for exclusion and inclusion of these measures. He stated that while the measures in discussion have been defined as "topped out" by CMS, NQF, or the MAP, the term "topped out" does not have a universal definition. He asked the TEP if, at the criteria level, there was strong feeling for or against including "topped out" measures in the summary star.</p> <p>One TEP member felt these measures should be included, but asked if the methods will allow high performers to be four or five star. Dr. Venkatesh explained that the third TEP meeting is designed to address how a summary score translates to a star rating. He said for the purposes of this TEP to think about "topped out" as high performance with little variation. The TEP member responded that if the measures continue to be active and are reported on <i>Hospital Compare</i>, they should be included.</p> <p>Another TEP member was concerned that the summary star rating should prioritize inclusion of measures for which there is meaningful</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>performance and allow information that is not included to be available to consumers on <i>Hospital Compare</i>.</p> <p>Dr. Venkatesh asked the TEP members if “topped out” measures are important to patients and consumers for inclusion in the summary star rating.</p> <p>One TEP member supported including the measures in star ratings as long as they are active on the CMS website and not retired by CMS.</p> <p>Another TEP member thought they should be excluded and that they may be low-bar measures, showing no strong correlation with outcomes.</p> <p>Another TEP member argued that these measures are reliable and are considered “topped out” for use in the Hospital Value-based Purchasing (HVBP) program. The TEP member added that if outliers are present, these measures are meaningful to the summary star ratings.</p> <p>One TEP member added that another publicly available star rating system does not include “topped out” measures because consumers tend to focus more on outcome measures “Topped out” may not be applicable to outcome measures for which typically a lower score indicates better quality.</p> <p>Another TEP member felt the “topped out” measure should not be easily dismissed. The TEP member stated that including these measures could serve as a useful educational tool for patients, helping them better communicate with their providers.</p> <p>Dr. Venkatesh acknowledged that there were varying opinions for including or excluding “topped out” measures in the summary star and stated that the CORE team will ensure that all input is available to CMS in addition to other public reporting constraints when considering this criteria.</p> <p>Dr. Venkatesh reviewed the survey results for efficiency measures and outlined reasons for exclusion and inclusion of these measures. He stated there are two non-directional measures in this group, OP-9 Mammography Follow-up Rates and MSPB-1/SPP-1 Medicare Spending per Beneficiary. The recommendation of the CORE team is that these measures should not be included in the summary star rating. He</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>provided reasons to include and exclude the other four imaging utilization measures.</p> <p>Dr. Venkatesh asked the TEP if there was any disagreement about exclusion of the two non-directional measures, and there was no response.</p> <p>Dr. Venkatesh asked the TEP for their thoughts on inclusion or exclusion of the imaging efficiency measures. There was discussion about potentially viewing an imaging efficiency measure as a proxy for patient safety. Dr. Venkatesh asked if there were any last comments around efficiency measures, and there was no response.</p> <p>Ms. Potteiger explained next steps, encouraged the TEP members to reach out to the CORE team with any questions, and thanked them for their participation.</p>	

TOPIC	NEXT STEPS	ACTION ITEM
<b>Next Steps</b>	<ul style="list-style-type: none"> <li>Two additional TEP meetings</li> <li>Two public comment periods</li> </ul>	(See action items above)

## Hospital Quality Star Ratings on Hospital Compare Technical Expert Panel

### Summary Call #2

Wednesday, February 4, 5:00-7:00 PM ET

#### PARTICIPANTS

**Technical Expert Panel (TEP):** Matt Austin, PhD; Vinita Bahl, DMD, MPP; John Bott, MBA, MS; Kathy Ciccone, RN, MBA; Rachel Grob, PhD; Rodney Hayward, MD; Emma Kopleff, MPH; Casey Schwarz, JD; David Shahian, MD; Brett Stauffer, MD, MHS; Guofen Yan, PhD

**Yale New Haven Health Services Corporation – Center for Outcomes Research (CORE):** Arjun Venkatesh, Jaymie Potteiger, Susannah Bernheim, Ben Clopper, Mallory Perez, Erica Norton, Zhenqiu Lin, Harlan Krumholz, Haiqun Lin, Angela Hsieh, and Nirupama Krishnamurthi

**Lantana Consulting Group (Lantana):** Kit Cooper, Hector Cariello

**National Opinion Research Center at the University of Chicago (NORC):** Romonda Bumpus

**Centers for Disease Control and Prevention (CDC):** Jonathan Edwards, Dawn Sievert

**Centers for Medicare & Medicaid Services (CMS):** Lein Han, Kristie Baus, Bill Lehrman, Karen Nakano

SUMMARY	ACTION ITEMS
<ul style="list-style-type: none"> <li>• Welcoming remarks and overview of project status</li> <li>• Measure Selection               <ul style="list-style-type: none"> <li>○ Review CMS decisions following TEP Meeting 1</li> <li>○ Review proposed exclusion criteria and measure list</li> </ul> </li> <li>• Summary Score Approaches               <ul style="list-style-type: none"> <li>○ Discuss standardization of measures and approaches for statistical model</li> <li>○ Discuss key decisions regarding latent variable modeling approach</li> </ul> </li> <li>• Next steps</li> </ul>	<ul style="list-style-type: none"> <li>• <b>CMS will share rationale and recommendations of the TEP with senior leadership to make key decisions.</b></li> <li>• <b>TEP members will provide CORE with recommendations for potentially expanding the scope of the star ratings.</b></li> <li>• <b>CORE will present the advantages and limitations of the proposed approach to hospital summary scores to CCSQ leadership.</b></li> <li>• <b>CORE will present options for grouping measures to CMS as they continue development of the star ratings.</b></li> <li>• <b>TEP members will reach out to CORE with any concerns regarding the proposed summary score approach.</b></li> </ul>

TOPIC	GENERAL
<p><b>Welcoming Remarks and Overview of Project Status</b></p>	<p>Jaymie Potteiger, MPH, Project Coordinator, CORE, welcomed participants to the second TEP meeting for the overall hospital star ratings on <i>Hospital Compare</i> project. Ms. Potteiger reviewed the meeting agenda, which included a project status overview, review of measure selection decisions, approaches to calculating summary score, and next steps.</p> <p>Ms. Potteiger mentioned that the overall hospital star ratings work is funded by a contract with the Centers for Medicare and Medicaid Services (CMS) Center for Clinical Standards and Quality (CCSQ). Additionally, she reminded participants that the discussion and materials from today's call are confidential and all project decisions and recommendations should remain confidential until publicly reported by CMS. She added that the summary report and other materials from the first TEP have been posted and are no longer considered confidential.</p> <p>Arjun Venkatesh, MD, MBA, MHS, Project Lead, CORE, reminded TEP members that the goal of the overall hospital star rating is to provide patients and consumers with information about multiple dimensions of quality in a single measure. He stated that in order to achieve the goal of improving the usability and interpretability of <i>Hospital Compare</i>, CMS will develop a methodology that can generate an overall star rating for each hospital using the existing measures available on <i>Hospital Compare</i>. He added that one of the inherent limitations to this work includes working with measures already in public reporting which may not fully capture all information related to a hospital's quality.</p> <p>Dr. Venkatesh reviewed the project timeline and stressed that it is very accelerated. He stated that the first public comment period is occurring in parallel with the second TEP and will close on February 25. A second public comment will also follow the third TEP meeting. CMS aims to have a preliminary methodology by April 2015 and a dry run over the summer, which would show hospitals their summary score derived from this methodology and encourage transparency. He explained that CMS may potentially report the star ratings on <i>Hospital Compare</i> but no formal decisions have been made. He stressed that the goal is to develop an approach that can be improved over time. He explained that the project has been divided into three phases:</p> <ul style="list-style-type: none"> <li>• Phase One: Measure selection</li> <li>• Phase Two: Hospital summary scores</li> <li>• Phase Three: Overall star rating</li> </ul> <p>Dr. Venkatesh discussed the objectives of today's meeting: to review the measure selection recommendations, to solicit feedback on a proposed approach for calculating hospital summary scores, and to discuss the key decisions for the proposed modeling approach.</p>

TOPIC	DISCUSSION	ACTION ITEM
<b>Measure Selection</b>	<p>Dr. Venkatesh summarized the measure selection discussion from the first TEP meeting. He reviewed CMS's decisions and proposed the following four exclusion criteria:</p> <ul style="list-style-type: none"> <li>• Measures suspended, retired, delayed, or awaiting public reporting on <i>Hospital Compare</i></li> <li>• Measures with less than or equal to 100 hospitals reporting</li> <li>• Structural measures without evidence of an association with changes in clinical practice or improved outcomes</li> <li>• Non-directional efficiency measures</li> </ul> <p>He added that the sample dataset used to conduct preliminary analyses includes 106 potential measures and after applying these exclusion criteria, 71 measures remain for inclusion in the overall star rating.</p> <p>One TEP member asked if any of the "topped out" measures were slated for exclusion per the current IPPS rule.</p> <p>Kristie Baus, RN, CMS, stated that the IPPS and OPSS rules for FY 2017 are in the process of being drafted and have not yet been determined for the upcoming year. The "topped out" measures excluded from the overall hospital star rating reflect CMS decisions included in the final FY 2016 IPPS rule.</p> <p>One TEP member commented on the importance of staying flexible while evolving as measurement evolves. The TEP member suggested separating the first exclusion criteria into two, measures removed from <i>Hospital Compare</i> and measures for future consideration.</p> <p>One TEP member asked if the TEP could discuss the implications of including "topped out" measures when reviewing the proposed latent variable approach.</p> <p>Dr. Venkatesh confirmed that the TEP member's question would be discussed when CORE reviews the proposed modeling approach.</p>	
<b>Summary Score Approaches</b>	<p>Dr. Venkatesh described the goals of an ideal model for generating hospital summary scores:</p> <ul style="list-style-type: none"> <li>• Generate a single aggregate measure of available hospital quality information</li> <li>• Limit subjective assumptions such as arbitrary waiting</li> </ul>	<b>CMS will share rationale and recommendations of the TEP with senior leadership</b>

TOPIC	DISCUSSION	ACTION ITEM
	<ul style="list-style-type: none"> <li>Accommodate changes in the included measures and hospital performance over time</li> <li>Use an approach that has been previously applied to similar work</li> </ul> <p>Dr. Venkatesh outlined some of the constraints to constructing hospital summary scores:</p> <ul style="list-style-type: none"> <li>Limited to existing publicly reported hospital quality measures</li> <li>Heterogeneity of measures available</li> <li>Each hospital reports a different number of measures</li> </ul> <p>Dr. Venkatesh stated that given these goals and constraints, CORE proposes using a latent variable modeling approach to generate the summary score for each hospital. He explained that this modeling approach assumes there is a latent quality trait that can be estimated from existing data.</p> <p>Dr. Venkatesh described the latent variable modeling approach graphic found on Slide 23 of the presentation. He stated that as new measures are introduced into public reporting and as hospital performance on measures changes over time, the “loadings” will evolve to be flexible to the available information. A hospital’s underlying quality (oval) influences its performance on each measure (rectangles), and this relationship is reflected through the measure’s loading (arrow thickness).</p> <p>Dr. Venkatesh outlined the standardization and Winsorization steps to the proposed process for generating summary scores. He explained that many of the measures selected for inclusion have different score formats and directions. Thus, the first step was to assign all measures to the same direction by converting all “lower score is better” measures into “higher score is better.”</p> <p>One TEP member mentioned that the latent variable analysis would reverse measures with a negative score naturally. The TEP member added that this additional step, while unnecessary, will not create any problems for calculating summary scores.</p> <p>Dr. Venkatesh noted that for the measures with an ordinal measure score format (the 11 HCAHPS measures), CORE used the continuous</p>	<p><b>to make key decisions.</b></p> <p><b>TEP members will provide CORE with recommendations for potentially expanding the scope of the star ratings.</b></p> <p><b>CORE will present the advantages and limitations of the proposed approach to hospital summary scores to CCSQ leadership.</b></p> <p><b>CORE will present options for grouping measures to CMS as they continue development of the star ratings.</b></p> <p><b>TEP members will reach out to CORE with any concerns regarding the proposed</b></p>

TOPIC	DISCUSSION	ACTION ITEM
	<p>HCAHPS scores that have been developed for HCAHPS Star Ratings. This will also align the data used in both Star Ratings.</p> <p>Dr. Venkatesh said that once all measures were transformed to a continuous measure score format and positive direction, CORE standardized the 71 measures to have a mean score of zero and a standard deviation of 1.</p> <p>Dr. Venkatesh explained Winsorization, the process for managing extreme outliers in the dataset whereby outlier values above the 99.5<sup>th</sup> percentile were set to the 99.5<sup>th</sup> percentile and values below the 0.5<sup>th</sup> percentile were set to the 0.5<sup>th</sup> percentile. CORE concluded that very few hospitals and very few measures were affected by Winsorization. The value of Winsorization is that it improves computation later in modeling.</p> <p>To orient the TEP to the statistical terminology used during the meeting, Dr. Venkatesh defined loadings as the magnitude of the relationship between the summary score and a given measure. Each measure's loading is the same for each hospital for the time the analysis is completed. He explained random effect as the statistical method to account for the relationship, or the correlation, between measures within hospitals. He also explained the concept of shrinkage, which allows the model to take hospitals with fewer available measures and provide an estimated summary score closer to the average of the distribution.</p> <p>One TEP member stated that shrinkage will have little effect on the model because the model aggregates measures. Shrinkage is of greater concern at the individual measure level.</p> <p>One TEP member asked if it is possible that more than one dimension of quality affects the summary scores when discussing the latent quality trait.</p> <p>Dr. Venkatesh stated the current latent variable modeling approach assumes quality can be measured as a single dimension. He added that a conceptual model for multiple dimensions of quality is likely true. CORE originally pursued this option but experienced conceptual and analytic challenges. He noted that every statistical approach makes assumptions and encouraged additional TEP feedback on the latent variable modeling approach.</p>	<p><b>summary score approach.</b></p>

TOPIC	DISCUSSION	ACTION ITEM
	<p>One TEP member stated that latent variable analysis is designed to test if there is a single dimension or multiple dimensions of quality.</p> <p>Bill Lehrman, PhD, CMS, mentioned that HCAHPS applied a different approach for small hospitals in which hospitals had to have a minimum number of cases for a measure before they were included in the star ratings. He asked if CORE was considering a similar approach.</p> <p>Dr. Venkatesh stated that CORE's general approach has been to include as much information as possible, including every reported hospital score when generating the model. However, CORE separates the decision to include measures in the hospital overall star rating from the decision to report a hospital's star rating on <i>Hospital Compare</i>.</p> <p>Susannah Bernheim, MHS, MD, CORE, added that the team has not yet made a decision about whether the five-star rating for hospitals with few measures will be reported.</p> <p>One TEP member asked if it is possible to do a missing data imputation to reduce shrinkage before applying a modeling approach.</p> <p>Dr. Venkatesh responded that previous feedback suggested that stakeholders do not like imputation and hospitals feel as though they are being assigned a score that lacks a meaningful application to their hospital.</p> <p>One TEP member was concerned that latent variable analysis may reward hospitals with good performance on process measures because process measures dominate the model. The TEP member argued that if this is the case, a hospital's summary score might not be indicative of their performance on outcome measures.</p> <p>Dr. Venkatesh explained that loadings are not weights, but rather loadings represent the summary score's relationship with a given measure. He agreed that if process measures in general have higher consistency, and therefore higher loadings, they would have more contribution to the summary score. However, he noted that higher loadings do not alone indicate that a hospital can have a high summary score with only high process measure performance.</p> <p>One TEP member stated that this is an important issue, but that latent variable analysis will empirically test if outcome measures aggregate differently from process measures.</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>Dr. Bernheim reiterated that a strength of this approach is that it is empirically derived. She explained that the data can indicate the relationship among the measures, which when combined reflect an unmeasurable concept of quality for each hospital.</p> <p>Dr. Venkatesh emphasized that a measure may have a high loading and that every hospital may do very well, and while this measure would be considered to have a strong relationship with latent quality, it may not be responsible for distinguishing one hospital's summary score from another hospital's.</p> <p>One TEP member questioned if the scenario presented remains true after standardization of the variances.</p> <p>One TEP member mentioned that lower loadings may also be very important and that the number of measures that reflect the same concept will influence hospitals' summary scores.</p> <p>Haiqun Lin, MD, PhD, CORE recognized that the number of measures reflecting the same concept influences hospitals' summary score. Dr. Lin explained that the proposed next steps are to explore whether process measures with high correlation could be combined into fewer measures.</p> <p>Dr. Venkatesh summarized the strengths of the latent variable modeling approach. This approach models a single underlying quality trait, accounts for the relationship between measures by predicting random-effect, and empirically derives loadings flexible to changes over time. He noted that some of the limitations are that the model is computationally challenging and the calculated loadings may not match traditional conceptual frameworks. He sought the TEP's input on the strengths and limitations of the latent variable modeling approach.</p> <p>Dr. Lehrman stated that the limitations do not seem to leave much room for CMS policy preferences. He asked if the model could incorporate policy preferences.</p> <p>Dr. Venkatesh responded that the model can incorporate policy preferences and CORE can test the impact of these preferences on hospitals' summary scores.</p> <p>One TEP member asked if the loadings were unique to a hospital.</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>CORE replied that the loading of a given measure is the same across all hospitals.</p> <p>A TEP member asked if there had been any consideration of incorporating confidence intervals into the summary score.</p> <p>CORE confirmed that the team has considered incorporating confidence intervals into its presentation of the summary scores to hospitals and for informing the assignment of star ratings.</p> <p>Another TEP member asked how the model assigns hospitals to one of five overall star rating categories.</p> <p>Dr. Venkatesh replied that translating summary scores to overall star ratings is the topic for the next TEP meeting.</p> <p>Dr. Venkatesh stated the first key decision is how to incorporate the HCAHPS measures of patient experience into the latent variable model. He said that the second key decision is how to approach individual measures that have negative loadings.</p> <p>Dr. Venkatesh stated that when running the latent variable model, the HCAHPS individual measures dominated the model. The HCAHPS measures are highly correlated with each other and the remainder of measures in the dataset; thus, the loadings of the HCAHPS measures are significantly higher than loadings of other measures. As a result, the hospital summary score would not be incrementally different from the HCAHPS summary star rating. Dr. Venkatesh added that a modification might be necessary to meet the original objectives of the Overall Hospital Star Ratings project.</p> <p>Dr. Venkatesh reviewed and explained the distribution of measure loadings when all HCAHPS measures are included in the model (N=71). He added that treating the correlation between the 11 HCAHPS measures the same as the correlation between other measures may not accurately portray the relationship between these measures. One modification for consideration is summarizing the 11 HCAHPS measures into one measure in alignment with the HCAHPS summary star rating. He stated the following advantages of summarizing the HCAHPS measures:</p> <ul style="list-style-type: none"> <li>• Maintaining measurement of patient experience in the overall hospital star rating</li> </ul>	

TOPIC	DISCUSSION	ACTION ITEM
	<ul style="list-style-type: none"> <li>• Approach to combining scores is well aligned with current HCAHPS star ratings</li> </ul> <p>He discussed the following disadvantages:</p> <ul style="list-style-type: none"> <li>• Inherent assumption that a single measure of patient experience captures contribution to latent “quality”</li> <li>• Inherent assumption of equally weighting each HCAHPS measure</li> </ul> <p>Dr. Venkatesh presented three options for handling the issue of HCAHPS measures having significantly higher loadings:</p> <ul style="list-style-type: none"> <li>• Make no modification, allowing HCAHPS measures to dominate;</li> <li>• Use the summary HCAHPS measure; or</li> <li>• Exclude HCAHPS from the overall star rating.</li> </ul> <p>One TEP member asked where the 11 HCAHPS measures come from and if CORE has considered including some number of HCAHPS measures between 11 and one in the model.</p> <p>Dr. Lehrman added that HCAHPS publicly reports on 11 HCAHPS measures, some of which are composites, and the method of creating the HCAHPS star rating will be publicly reported in April.</p> <p>Dr. Lehrman suggested that process and outcome measures be rolled up in a similar fashion as HCAHPS measures if the decision is to use the HCAHPS summary measure in the model. He added there is a lot more variability in hospital performance across the HCAHPS measures and suggested this might explain their high loadings.</p> <p>One TEP member stated this would not have any impact on measure loadings because variances are standardized. The TEP member added that the high loadings are attributable to the higher correlation between satisfaction and patient experience measures. The latent variable model uses a weighting of reliability, not weighting of importance. For this reason, the TEP member advocated for consideration of the patient-reported experience measures as a separate domain.</p> <p>Dr. Lehrman added that patient experience measures are treated as a separate domain in Hospital Value-Based Purchasing.</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>Dr. Venkatesh outlined the two reasons why domain scores may not be feasible at this time. Domain scores require assigning an arbitrary weight between the domains, which may not allow for flexibility over time or reflect each stakeholder's relative importance of that domain. Additionally, hospitals with a different number of measures available may lead to missing domains. He sought the TEP's feedback on domain scores.</p> <p>One TEP member voiced concern that if the overall hospital star rating combines patient experience measures and process measures, the information conveyed by these measures is reduced. The TEP member asked if the model will reflect distinctions between measures of differing measure types if domain scores are not used.</p> <p>Dr. Lin replied that the software does not naturally distinguish between measure type.</p> <p>Dr. Venkatesh added that the general assumption is that we are limited to the existing measures on <i>Hospital Compare</i>.</p> <p>A TEP member asked if the need for a single domain of star ratings is being externally driven by CMS or if it is CORE's preference.</p> <p>Lein Han, PhD, CMS, stated that CMS is considering implementing the star ratings because CMS believes it would be helpful to the patient to comprehensively interpret and understand these 71 measures. Dr. Han said that they would take the rationale and recommendations of the TEP to their senior leadership.</p> <p>Another TEP member stated that these 71 measures should not be combined because they represent different factors of quality. The TEP member did not agree that a single star rating will be as valuable to patients as domain stars</p> <p>One TEP member suggested letting the data determine how many latent variables exist among the 71 measures.</p> <p>Dr. Han stated that CMS has interest in domains for the future.</p> <p>Ms. Baus added that CMS's goal is to come up with one overall star rating based on the results of consumer testing done a few years ago. She added that CMS also wants to have domain scores, but CMS is currently grappling with the best way to determine the domains and assign measures.</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>Dr. Bernheim stated that when the overall star ratings project first started, CORE considered creating domain stars, but for the reasons that were outlined earlier, CORE decided to conceptualize a single model that generated an overall score first. She said the focus of the next TEP meeting is to determine how to take a single summary score and quantify it into stars. CORE may consider separate domain ratings after proposing an overall star rating methodology. She asked the TEP to give feedback about their preferences for domains.</p> <p>Several TEP members stated a preference for starting with the overall score and working towards the eventual ability to drill down into the domains. Several TEP members stated the need to respect what the consumers have identified as being most useful or important to them.</p> <p>One TEP member voiced concern that aggregating measures shown to be empirically different would not produce a useful and meaningful measure for the consumer.</p> <p>Dr. Venkatesh said that as CORE was initially developing this model, they tried to allow the data to define several latent variables through a form of factor analysis. The model did not converge using factor analysis because of missing data and the variability in the number of available measures across hospitals. He added that if the TEP is interested in domains, those domains would have to be conceptually defined and would require the process of assigning measures to them. He sought the TEP's feedback on domain stars and asked for thoughts about CORE's proposal to summarize the HCAHPS measures for the overall hospital star rating.</p> <p>Dr. Venkatesh also asked the TEP for feedback on the importance of domain stars "rolling up" to the overall hospital star rating. He sought input from the TEP, hypothesizing if CORE expanded the scope of the project to generate domain stars, would the TEP recommend that CMS develop the overall star rating and then develop a methodology for domain stars that may not roll up to the overall star rating.</p> <p>One TEP member inquired about how CMS plans to portray the data to the public.</p> <p>Another TEP member supported proceeding with a single overall star rating using latent variable modeling.</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>Dr. Lehrman asked if it is possible to create three domains from the HCAHPS data rather than one.</p> <p>Dr. Venkatesh stated that this could be done, but the question of weighting remains.</p> <p>Dr. Lehrman said that weighting becomes arbitrary and these might be options to present to CCSQ leadership.</p> <p>One TEP member was concerned that the correlation of the HCAHPS measures may also exist for other measures, and, if so, weighting measures should not be arbitrary.</p> <p>Dr. Bernheim summarized the TEP's feedback. The TEP generally approved of the latent variable modeling approach, and there was interest in thinking about ways to either separate measures into different domains or create some additional summary measures. She added her impression that the TEP encouraged CORE to be creative in finding the best ways to combine and separate measures. She stated CORE would bring this work back to CMS.</p> <p>One TEP member stated there are some potential problems with the latent variable approach in that it assumes that the amount of variance is an important issue in capturing some domain. The TEP member specifically referenced "topped out" measures because if variance is small, quality is a smaller component of the residual variance.</p> <p>Dr. Venkatesh confirmed the TEP member's feedback and stated that CORE will reflect on this concern when preparing documents for public comment.</p> <p>Dr. Venkatesh discussed the issue of measures with negative loadings. He asked for feedback from the group about whether or not to include these measures in the summary score.</p> <p>One TEP member did not agree with keeping measures with negative loadings because it penalizes those hospitals that are actually doing better and suggested dropping these measures if the variation is minimal.</p> <p>Angela Hsieh, PhD, CORE responded that the correlation between summary scores and negative measures is very low.</p> <p>Dr. Venkatesh stated that the low correlation of these measures led CORE to conclude that there is minimal impact between these</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>measures and the summary score. He added that loadings would evolve over time as hospital performance and the number of measures included in the model changes.</p> <p>Dr. Venkatesh thanked the TEP members for their time and encouraged all individuals to reach out to CORE with any additional feedback or comments.</p>	

TOPIC	NEXT STEPS	ACTION ITEM
<b>Next Steps</b>	<ul style="list-style-type: none"> <li>Review next analytic steps</li> <li>Review next project steps</li> </ul>	

## Hospital Quality Star Ratings on Hospital Compare

### Technical Expert Panel

#### Summary Call #3

Monday, March 30, 4:00-6:00 PM ET

#### PARTICIPANTS

**Technical Expert Panel (TEP):** Matt Austin, PhD; Vinita Bahl, DMD, MPP; John Bott, MSSW, MBA; Kathy Ciccone, RN, MBA; Kelly Court, RN, MBA; Rachel Grob, PhD; Rodney Hayward, MD; Emma Kopleff, MPH, Laura Petersen, MD, MPH, FACP; Casey Schwarz, JD; David Shahian, MD; Guofen Yan, PhD; Ben Yandell, PhD

**Yale New Haven Health Services Corporation – Center for Outcomes Research (CORE):** Arjun Venkatesh, Jaymie Potteiger, Susannah Bernheim, Haiqun Lin, Ben Clopper, Jeph Herrin, Angela Hsieh, Nirupama Krishnamurthi, Zhenqiu Lin, Mallory Perez, Erica Norton, Elizabeth Drye

**Lantana Consulting Group (Lantana):** Kit Cooper

**Centers for Disease Control and Prevention (CDC):** Jonathan Edwards, Dawn Sievert, Lindsey Weiner

**Centers for Medicare & Medicaid Services (CMS):** Bill Lehrman, Liz Goldstein, Karen Nakano, Gray Sim, Lein Han, Kia Stanfield

SUMMARY	ACTION ITEMS
<ul style="list-style-type: none"> <li>• Welcoming remarks and overview of project status</li> <li>• Public Comment 1 summary <ul style="list-style-type: none"> <li>◦ Reviewed general and criteria-specific comments received during the first public comment period</li> </ul> </li> <li>• Proposed approach to generating hospital summary scores <ul style="list-style-type: none"> <li>◦ Discussed the two-stage approach using latent variable modeling (LVM) and a weighted average of measure group summary scores.</li> <li>◦ Solicited feedback on the proposed weighting scheme for the measure groups and the relative importance of each group.</li> </ul> </li> <li>• Options for translating summary scores to star ratings <ul style="list-style-type: none"> <li>◦ Reviewed and discussed three proposed options for translating the summary scores to star ratings with most feedback supporting Options 2 and 3.</li> </ul> </li> <li>• Missing measures and domains <ul style="list-style-type: none"> <li>◦ Discussed options for setting thresholds hospitals must meet to have a star rating publicly reported on <i>Hospital Compare</i>.</li> </ul> </li> <li>• Next steps</li> </ul>	<ul style="list-style-type: none"> <li>• <b>CMS will take rationale and recommendations of the TEP to senior CMS leadership.</b></li> <li>• <b>CORE team will discuss with CMS the removal/addition of groups.</b></li> <li>• <b>CORE team will discuss with CMS the appropriate mechanism for determining measures removed from <i>Hospital Compare</i>.</b></li> <li>• <b>TEP members may submit any additional questions or feedback to the CMS Star Ratings inbox (cmsstarratings@yale.edu).</b></li> </ul>

TOPIC	GENERAL
<p><b>Welcoming Remarks and Overview of Project Status</b></p>	<p>Jaymie Potteiger, MPH, Project Coordinator, CORE, welcomed participants to the third TEP meeting for the Overall Hospital Quality Star Ratings on the <i>Hospital Compare</i> project. Ms. Potteiger reviewed the meeting agenda, which included an overview of project progress, a summary of the first public comment period, the proposed approach to generating hospital summary scores, options for translating summary scores to star ratings, the potential approaches for missing measures and groups, and next steps.</p> <p>Ms. Potteiger mentioned that this work is funded by a contract with the CMS Center for Clinical Standards and Quality (CCSQ). She reminded participants that the discussion and materials from today's call are confidential and all project decisions and recommendations should remain confidential until publicly reported by CMS. She added that this does not include the first public comment period materials, which have been publicly posted and are no longer considered confidential.</p> <p>Arjun Venkatesh, MD, MBA, MHS, Project Lead, CORE, thanked the attendees for their participation in the TEP. He reminded TEP members that the goal of the Overall Hospital Quality Star Ratings is to provide patients and consumers with data that will inform them about multiple dimensions of quality in a single measure. He requested the TEP keep this goal in mind as they discuss a variety of decisions. He emphasized that the project goal is to improve the usability and interpretability of <i>Hospital Compare</i> for patients and consumers and to develop a methodology that can generate an overall star rating using existing measures available on <i>Hospital Compare</i>.</p> <p>Dr. Venkatesh acknowledged the limitations of this work and stressed the importance of transparency with the TEP members and other key stakeholders. He discussed the following limitations: the existing measures on <i>Hospital Compare</i> do not capture all aspects of hospital quality, are of varying measure types, and that this project seeks to combine distinct elements of quality into a single measure while recognizing that not all hospitals of the same rating are the same. He acknowledged that there is no perfect solution; however, the recommended approach strives to make the data more usable to consumers and patients.</p> <p>Dr. Venkatesh reviewed the objectives of the meeting, starting with a review of the summary of comments received during the first public comment. Second, the TEP will review feedback received from internal experts at CMS regarding the proposed modeling approach and weighting scheme. Next, CORE will seek TEP input on the options for translating hospital summary scores to star ratings. The TEP will also discuss options for handling instances of missing measures and missing measure groups. The meeting will end with a review of next steps.</p>

TOPIC	GENERAL
	<p>Dr. Venkatesh reviewed the project timeline. In the spring, CMS will hold a second public comment period presenting the proposed methodology, approach for translation to stars, and summaries of TEP discussions from the second and third TEP meetings. He noted that in parallel to the second public comment period, there will be a dry run to provide hospitals with hospital-specific reports specific to Overall Hospital Quality Star Ratings and elicit hospital feedback. This will allow for a wide variety of input from stakeholders prior to finalizing a proposed methodology.</p> <p>Dr. Venkatesh reviewed the status of the three phases of the project. He noted in Phase 1, CMS considered the TEP's feedback to narrow the list of measures for inclusion in the Overall Hospital Quality Star Ratings. The proposed measure exclusion criteria were presented during the public comment, and 12 unique responses were received. In Phase 2 of the project, the TEP will review the proposed two-stage approach for calculating hospital summary scores. Phase 3 involves three potential approaches for categorizing hospitals into star ratings.</p>

TOPIC	DISCUSSION	ACTION ITEM
<b>Review Public Comment Summary</b>	<p>Dr. Venkatesh reviewed the feedback received from the first public comment period. He noted that the 12 commenters were mostly hospitals and providers and there was no public comment from consumers or advocacy groups. The following bullets summarize the comments with general feedback.</p> <ul style="list-style-type: none"> <li>• Some commenters argued against a single summary star rating and recommended CMS consider domain or topical star ratings.</li> <li>• Two commenters questioned the decision to include Outpatient Quality Reporting (OQR) program measures because these measures are voluntarily reported by hospitals and therefore may not be valid or reliable.</li> <li>• A number of comments call into question both the lack of risk adjustment in some outcome measures (e.g., SDS) and adequacy of the adjustment of other measures.</li> <li>• Some commenters expressed concerns that multiple varying rating methodologies available to the public can be confusing and misleading.</li> </ul> <p>Dr. Venkatesh said that CORE brought these comments to CMS. The comments reveal a variety of topics that will need support with</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>communication materials. He noted that these comments would inform future communication efforts.</p> <p>Dr. Venkatesh reviewed the specific comments received from the public comment regarding the criteria proposed for including and excluding measures.</p> <ul style="list-style-type: none"> <li>• Commenters generally agreed with the criteria to exclude measures with less than 100 hospitals publicly reporting performance.</li> <li>• Commenters largely supported the exclusion of non-directional efficiency measures, but were split on the inclusion of directional efficiency measures, which are currently included in the star ratings.</li> <li>• Commenters had mixed feedback regarding the inclusion of “topped out” measures, which are currently included in the star ratings.</li> <li>• Commenters did not favor the current decision to include measures that have been de-endorsed by the National Quality Forum (NQF) and not supported by the Measure Applications Partnership (MAP).</li> </ul> <p>Dr. Venkatesh noted that the final comment was made at the criterion level and did not consider the timeline implications discussed during TEP Meeting 1. He said that the CMS decision to include NQF de-endorsed measures was based on TEP feedback to ensure that Overall Hospital Quality Star Ratings remain consistent with <i>Hospital Compare</i> and include as many publicly reported measures as possible. He noted that the de-endorsed measures are often removed from <i>Hospital Compare</i> and would therefore be removed from star ratings in subsequent updates.</p>	
<b>Proposed Approach to Generating Hospital Summary Scores</b>	<p>Dr. Venkatesh outlined the goals of the proposed approach for generating hospital summary scores:</p> <ul style="list-style-type: none"> <li>• Produce a single summary score</li> <li>• Incorporate diverse data while accounting for missing measures and groups</li> <li>• Capture the multidimensionality of the data</li> <li>• Update the data every quarter with new measures being added or removed</li> </ul>	<b>CMS will take the rationale and recommendations of the TEP to senior leadership.</b>

TOPIC	DISCUSSION	ACTION ITEM
	<ul style="list-style-type: none"> <li>Allow for policy-driven weighting versus a purely data-driven approach</li> </ul> <p>Dr. Venkatesh summarized the past feedback received from the TEP and other statisticians regarding the approach to generating hospital summary scores. He noted the desire expressed by stakeholders for the approach to reflect the multidimensional nature of the quality data. He stressed that latent variable modeling (LVM) assumes that consistent measures carry a stronger signal towards the latent trait. He noted that some stakeholders agree with this assumption while others disagree. He also noted that LVM might be difficult for patients and consumers to understand, as this modeling technique is not currently used in other CMS star ratings efforts. Lastly, stakeholders recommended the model account for the sampling variance. He noted CORE is working on methods to account for this in the approach.</p> <p>Dr. Venkatesh noted that the proposed two-stage approach addresses many of the concerns reflected in this feedback.</p> <p>Dr. Venkatesh described the two-stage approach, noting that in the first stage, LVM is applied to each measure group, resulting in six separate latent variable models. He noted this stage acknowledges the multidimensionality of hospital quality. The second stage of the model combines the group-specific scores into an overall hospital summary score using a policy-driven weighting scheme.</p> <p>Dr. Venkatesh noted the current six measure groups are patient experience measures (currently comprised of HCAHPS measures), outcome – safety measures (including HAI measures), outcome – mortality measures, outcome – readmission measures, process measures, and efficiency measures (outpatient imaging measures). He said this grouping method could be reliably applied to current and future measures. He noted that the process measures group is the largest, and therefore, a little more heterogeneous.</p> <p>Dr. Venkatesh reviewed the path diagrams demonstrating how the measures will be grouped and how the group-specific scores will be used to calculate the summary score. He noted that each measure group was based on existing clinical and policy groupings and that this approach would generate six scores reflecting different aspects of quality, each with a latent group summary score. He continued that</p>	<p><b>CORE team will discuss the removal/addition of groups with CMS.</b></p> <p><b>CORE team will discuss with CMS the appropriate mechanism for determining measures removed from <i>Hospital Compare</i>.</b></p>

TOPIC	DISCUSSION	ACTION ITEM
	<p>these scores would be combined using a simple weighted average based on the policy-driven weights.</p> <p>Dr. Venkatesh reviewed the advantages and disadvantages of the two-stage approach. He noted that the advantages include:</p> <ul style="list-style-type: none"> <li>• The method is reproducible and feasible in comparison to the weighted average of individual measures to generate a group score.</li> <li>• Loadings account for consistency or correlation between measures within a group.</li> <li>• LVM allows for missing measures by using only available information to generate a group score. Hospitals with few measures are pulled towards the mean.</li> <li>• Sampling variance can be accounted for within the model.</li> <li>• Confidence intervals can be calculated around the overall summary scores.</li> <li>• The method is used for composite measures in health quality literature.</li> </ul> <p>Dr. Venkatesh reviewed the limitations of the two-stage approach and noted it is important to be transparent about these elements:</p> <ul style="list-style-type: none"> <li>• The overall summary score requires a subjective, policy-based weighting scheme.</li> <li>• The use of a LVM for the patient experience domain could result in similar but different patient experience summary scores than HCAHPS Star Ratings.</li> <li>• The model complexity may be challenging for consumers and patients to understand, thus requiring effective communication.</li> <li>• Other CMS star ratings efforts use different methodologies for combining measures potentially leading to differences in rankings.</li> </ul> <p>Dr. Venkatesh described the diagrams found on Slides 22 and 23 of the presentation, showing the loadings for each of the six measure groups. He noted that in general, the groupings used for the outcome measures make sense clinically and the loadings support using LVM. He reminded the group that these loadings will be recalculated every three months and, therefore, will change as the data evolves.</p> <p>One TEP member asked if measure loadings would be publicly reported and updated every three months. Dr. Venkatesh responded that this topic is still in discussion among CMS and asked if the TEP member has a recommendation for CMS. The TEP member urged for maximum</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>transparency and noted that even if the burden would be high, it would be important for hospitals to understand their performance and for consumers to understand how the ratings were constructed.</p> <p>One TEP member presented concerns that other than the process measures and the satisfaction measures, this model is sending a strong message that will hurt patients. The TEP member noted that the single star rating as proposed endorses the measures with poor risk adjustment for social circumstances, which may cause more problems. The TEP member stated that other than the process measures and the satisfaction measures, the included measures do not capture quality to a high degree. The TEP member requested that these other measure groups not be over-weighted since they have less-than-optimal science behind them.</p> <p>Susannah Bernheim, MD, MHS, CORE, responded that there is a wide range of opinions on the validity of the measures. She noted that risk adjustment using sociodemographic variables is a very active discussion within CMS and NQF and that as the science evolves, the measures will evolve as well. She stated that there is a thorough process for vetting measures prior to adding them to <i>Hospital Compare</i> and we are trying to find the best way to reflect what is already approved for reporting.</p> <p>Dr. Venkatesh reviewed the diagram on Slide 23 of the presentation, which demonstrates the efficiency and process group loadings. He noted that the efficiency grouping, for which CORE is still in the process of accounting for sampling variance, does not fit as well as the outcome and patient experience groups with the latent variable model. He stated, however, that the fit is reasonable.</p> <p>Dr. Venkatesh noted that the process measure group had three measures with negative loadings, keeping in mind that the group is more heterogeneous. He requested suggestions from the TEP on alternative ways to group these measures.</p> <p>One TEP member noted that the level of correlation between diseases is modest in the process measures and therefore, the area with more measures will dominate the latent variable analysis. He said that this is one of the limitations of LVM that will need to be addressed.</p> <p>One TEP member asked how the addition or removal of measure groups would be accommodated under the proposed approach. Dr.</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>Venkatesh noted that this is a good question and it will be taken back to CMS for guidance. He noted that the proposed six groups were selected for this potential version of the star ratings. In the future, the groups could be revised if new or different groups are more appropriate.</p> <p>One TEP member asked if an analysis had been done to separate the process measures into several groups. Dr. Venkatesh noted that this analysis is in progress and is challenging since the earlier analysis indicated that conceptually similar measures such as stroke process measures do not map to a common empirical construct. The TEP member suggested grouping measures by care setting (i.e., outpatient and inpatient settings). Dr. Venkatesh stated that CORE would explore this suggestion.</p> <p>One TEP member noted that data collection has stopped since December 2014 for at least half of the process measures presented on Slide 23. The TEP member inquired about what the process group would look like with those measures removed. Dr. Venkatesh questioned if the measures the TEP member was referring to are outside of the exclusion criterion of retired measures and are instead in a limbo state of being reported on <i>Hospital Compare</i>. The TEP member confirmed this and noted that these measures are scheduled for removal and suggested they be removed from the current modeling.</p> <p>Bill Lehrman, PhD, CMS, recommended cross-referencing all of the groups' measures to see which are outpatient and which are inpatient, as other outpatient measures could also be moved to a new group.</p> <p>One TEP member recommended that before removing the measures that were retired in December 2014, CORE should review the reason for retirement. Dr. Venkatesh noted that CORE would research how to incorporate future retired and added measures into the dry run and for the second public comment.</p> <p>Dr. Venkatesh noted that there seem to be two discrepancies; the first is with respect to measures put forth in front of the public, and the second is the measures that are slated for future removal from <i>Hospital Compare</i> but are still currently reported. He noted CORE would bring these issues back to CMS for consideration.</p> <p>Dr. Venkatesh presented the topic of weighting for discussion, requesting TEP feedback on two decisions: how to prioritize measure</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>groups and how to handle missing groups. He noted that CORE considered two options for weighting the groups (equal weighting and a weighting scheme modified from hospital value-based purchasing (HVBP)). He noted the proposed HVBP-modified weights are illustrated on Slide 27 of the presentation and the third column is just a starting point for discussion.</p> <p>Dr. Venkatesh noted there are two fundamental differences between the HVBP program and star ratings. First, readmission measures are not part of HVBP because they are a part of a separate CMS program. Second, HVBP includes only the Medicare Spending per Beneficiary measure in its efficiency domain, which is not included in star ratings.</p> <p>Dr. Venkatesh noted that to help determine the relative importance of each group, he would first review the responses from the pre-TEP survey. The results represent the mean score of relative importance for 13 respondents; each group was rated between 1 and 10 with 10 being the most important to patients and consumers relative to the other groups. He noted that the outcome-mortality and outcome-safety groups were ranked as most important. He also noted that the efficiency and process groups had the highest number of low ranks. He stated that the ranking of the HCAHPS and outcome-readmissions groups were less consistent.</p> <p>One TEP member noted that the proposed weights make sense when you account for the importance of each domain, but makes little sense if you view it from the science behind the current measures in each domain. The TEP member noted that the process measures have the best science, citing researcher Jonathan Mann.</p> <p>Dr. Venkatesh asked if the TEP member recommends higher weights for the process and patient experience domains. The TEP member recommended that the patient experience domain be a separate star rating from technical quality; however, the process measures are the best measures available. The TEP member noted the other measures could be fixed theoretically, but have not been fixed yet. The TEP member is concerned that highly weighting measures with inadequate risk-adjustment will result in gaming that could ultimately be more harmful to the public.</p> <p>One TEP member agreed that the process measures are the best measures and are probably getting a low rating with HVBP because</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>there are only three measures in the program. The TEP member argued that the Patient Experience measures should be included in the star ratings as they have an aspect of quality not picked up in some of the clinical quality measures. The TEP member also noted that the outcome-safety group includes some measures based on CMS claims data and other measures using data from the National Healthcare Safety Network (NHSN). The TEP member stated that all measures suffer from a problem of inadequate risk adjustment.</p> <p>Dr. Bernheim clarified that this project is not about changing measures' methodologies and acknowledged that the measure science will continue to evolve. She noted the purpose of the TEP is to provide input on the relative importance of these measures to stakeholders.</p> <p>One TEP member noted that the weighting should be similar to HVBP, as it will cause confusion and mixed signals if the weighting is not aligned. The TEP member said the biggest thing that member hospitals complain about is misalignment between CMS programs.</p> <p>One TEP member agreed with the proposed weighting. The TEP member also urged the need for transparency and the ability for hospitals to reproduce their scores. The TEP member concurred with the idea of harmonization between CMS programs.</p> <p>Dr. Venkatesh responded that transparency could be enhanced in several ways, including having preview reports and providing as much detail as possible in technical documentation. He noted a challenge is that a hospital's performance is largely based on its comparison to that of other hospitals. He said therefore, in the absence of having every hospital's data, a hospital cannot reproduce their score. He noted that information could be provided on which measures were included and how these measures were grouped.</p> <p>One TEP member noted that the intent of this work is to get consumers what they want, and therefore we should follow the literature that shows consumers prefer outcome measures. The TEP member noted that while there are weaknesses, NQF did endorse the measures. The TEP member also noted that in several studies, process measures are poorly correlated with outcomes. The TEP member stated that the TEP should also look at where CMS is putting its dollars and that the vast majority of CMS's dollars is going towards readmissions.</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>One TEP member responded that there is concern around putting a lot of weight towards readmissions since many readmissions have to do with the locale and the patient population. The readmission measures have not been able to incorporate the needed socioeconomic risk–adjustment variables. The TEP member recommended a lower weight for readmission measures.</p> <p>One TEP member agreed with the primacy of outcomes, especially for mortality. The TEP member added that certain safety measures are not valid, but the validity of the readmission measures is improving. The TEP member argued that process measures should not be given more weight as they do not correlate with outcome measures based on available literature.</p> <p>One TEP member suggested grounding the weighting scheme in a theory that will help inform weighting in the future, rebalancing the weights as domains come and go. Dr. Venkatesh responded that an environmental study was completed for similar work. He noted that CORE proposed HVBP-modified weights because HVBP has similar measure groups to star ratings, and, in HVBP as well as the CMS Quality Strategy, outcome measures have more weight.</p> <p>One TEP member noted that the poor correlation between process and outcome measures is as to be expected. The TEP member added that factors outside of a hospital’s control influence the hospital’s performance on outcome measures. Conversely, the TEP member argued that process measures only account for what is within the hospital’s control.</p>	
<b>Translating Summary Scores to Star Ratings</b>	<p>Dr. Venkatesh reviewed the initial assumptions for translating summary scores into star ratings. He noted the first assumption is that the star ratings will be whole stars from one to five (no half stars). He noted the second assumption is that a hospital’s star rating is a reflection of the available quality information for the hospital at that given time. He noted the third assumption is that three stars would be considered “average.” He stated the final assumption is that supplementary information will be provided on the website to improve understandability.</p> <p>Dr. Venkatesh proposed three options for the star ratings calculations. He noted that a decision to select one of the three options would need</p>	<b>CMS will take the rationale and recommendations of the TEP to senior leadership.</b>

TOPIC	DISCUSSION	ACTION ITEM
	<p>to be based on policy considerations as each option has technical complexities.</p> <p>Dr. Venkatesh explained that Option 1 (located on Slide 33 of the presentation) would use a fixed interval approach where hospitals are broken up into quintiles. He noted advantages of this option included the ability to have specific proportions of hospitals in each category and that the approach could be simpler to explain to hospitals and the public. He cautioned that the limitation of the option is that there would be no meaningful difference between the edges of the categories (e.g., 20<sup>th</sup> versus 21<sup>st</sup> percentile).</p> <p>Dr. Venkatesh explained that Option 2 (located on Slide 36 of the presentation) sets an overall threshold for how a hospital's summary score compares to other hospitals and then sets a threshold along individual measures. He provided an example of this, as a five-star would be significantly better than the national average and more than half of the individual measures are better than the national average. He noted the advantages of this option are that the categories reflect statistical differences in the summary scores and hospitals with only a few measures must perform well on the same percentage of total measures reported as hospitals with many measures. He highlighted the following limitations of this option: the frequency of the hospital in each category will likely be unequal, the majority of hospitals fall into the three-star category (92.13 percent based on test data), it requires face validity to set the threshold, and it may appear to "double count" individual measure performance.</p> <p>Dr. Venkatesh explained that Option 3 (located on Slide 39 of the presentation) is the use of <i>k</i>-means clustering analysis. He noted this is based on the idea that there are centers of gravity amongst the distribution of the hospital summary scores and there is a way to statistically group or bucket each of these hospitals to their closest center of gravity. He said the number of buckets is fixed by the model, so the current option has selected five buckets for the five stars. He noted that the five buckets from this model are located on Slide 41 of the presentation. He noted the advantages of this option are that hospitals in a cluster have statistically similar summary scores and individual measure performance does not appear to be assessed "twice" as in the threshold approach. He cautioned that the limitations</p>	

TOPIC	DISCUSSION	ACTION ITEM
	<p>of this option include that the majority of hospitals fall into the three-star cluster and the complexity of the approach may be difficult for patients and consumers to understand.</p> <p>Dr. Lehrman noted that <i>k</i>-means clustering is being considered for HCAHPS and for other star ratings efforts.</p> <p>One TEP member noted that the <i>k</i>-means clustering option is the most attractive, while Option 1 (fixed intervals) creates arbitrary break points that may not be meaningful. The member added that the <i>k</i>-means clustering option provides better separation than the second option.</p> <p>One TEP member agreed that either the threshold or the <i>k</i>-means clustering options would be intuitive approaches for beneficiaries. The TEP member noted that it is important to include the rationale in the public materials. The TEP member agreed that the interval option is awkward.</p> <p>One TEP member agreed with Option 2 and Option 3 but noted that the distributions shown on the presentation will not necessarily be the actual distributions since the parameters can be changed. The TEP member cautioned that the situation would arise of two hospitals being very close in terms of summary scores but ending up in different buckets.</p> <p>One TEP member asked if a comparison has been made between the classification of a hospital and the three approaches (e.g., what is one hospital's score under Option 1 versus under Option 3). Dr. Venkatesh noted this has not been completed with the two-stage approach but that the CORE team intends to complete this analysis in the near future.</p> <p>The TEP member asked if sampling variation is taken into account for a single summary score in terms of a confidence interval. Angela Hsieh, PhD, Lead Analyst, CORE, noted that the sampling variation is included in Option 1 of the approach and the confidence interval is used in Option 2 but not in Option 3.</p> <p>One TEP member noted that hospitals should be incentivized to do well and therefore the approach should not set a cap for the top tier.</p> <p>Dr. Venkatesh asked the group to also consider a potential situation where Option 2 would lead to many more three-star hospitals than Option 3.</p>	

TOPIC	DISCUSSION	ACTION ITEM
	One TEP member noted that one of the appealing aspects of Option 1 was that there is less of a problem with clustering performance around three stars.	
<b>Missing Measures and Domains</b>	<p>Dr. Venkatesh noted that one of the challenges of using the two-stage approach with domains is how to determine when there is enough information about a hospital to assign a star rating. He noted that all analyses presented to the group have included every hospital with at least one measure. He explained it was necessary to determine a threshold for both the minimum number of measures required for calculating a measure type score, but also a minimum number of measure type groups to calculate the hospital summary score.</p> <p>Dr. Venkatesh reported that about 4,700 hospitals have at least one measure and the median number of measures hospitals have is 51 measures. He cautioned there is a wide variation between the percentiles, where 25% of hospitals have less than 21 measures and the 75<sup>th</sup> percentile has 64 measures. He noted that Slide 45 in the presentation shows these numbers based on the proposed measure type groups. He said over 3,000 hospitals have at least one measure across all six groups; however, there are almost 500 hospitals that may only have three or four groups of information available.</p> <p>Dr. Venkatesh noted Slide 46 of the presentation illustrates the trade-off of hospitals based on the thresholds imposed. He stated the thresholds provided range from very liberal, with two measures available in at least three groups (87% of hospitals), to very strict thresholds, with at least four measures in all six groups (40% of hospitals). He inquired whether anyone had feedback on what these requirements should be.</p> <p>One TEP member noted that if a hospital does not have enough measures in one measure group, then the weighting is useless since only measure groups with measures end up receiving weight and, therefore, skew the results.</p> <p>One TEP member noted that since some measure groups are weighted higher than other groups, hospitals missing groups that are considered more important should not display a star rating. The TEP member continued to say that for the process measures or efficiency measures that have a lower weighting, a hospital missing these groups could have</p>	<b>CMS will take the rationale and recommendations of the TEP to senior leadership.</b>

TOPIC	DISCUSSION	ACTION ITEM
	<p>the 5 percent redistributed. The TEP member asked for clarification on how the summary star calculation handles missing data.</p> <p>Dr. Hsieh noted that missing values do not contribute to the likelihood in the model. Dr. Venkatesh responded that in a few current programs when a group is missing, there is a re-proportional distribution across the remainder of the groups. He also clarified that the table on Slide 46 would require that at least one of the outcome groups be available. He noted that with the weighting method, if a re-proportioning of weights process were not used, it would be difficult to resolve missing groups short of requiring all domains to be available.</p> <p>Dr. Bernheim reminded the group that the proposed weighting scale presented is not final, but acknowledged the TEP member's comment on handling missing groups differently depending on the group's weight.</p> <p>One TEP member suggested using a percentage versus a fixed number, providing the example of at least half the measures in the group. The TEP member noted that otherwise, groups that have only a few measures would be weighted unfairly and hospitals might end up with a star rating based on only a few measures. Dr. Venkatesh noted that the team would explore this option further.</p> <p>One TEP member suggested that the developers consider that if there are key measures within a group that indicate more influence on the measure group score, that could be used as the minimum number of measures. The TEP member also added that even if a hospital does not have the minimum number of groups, they should have a measure type score reported.</p> <p>Dr. Venkatesh reminded the group that all measures included in star ratings would still be on <i>Hospital Compare</i>. He noted that CORE will take the idea of reporting measure group scores back to CMS, but the current work is limited to just the overall star rating.</p> <p>One TEP member suggested using the Donabedian model of outcome, process, structure, efficiency, and patient experience as the groups for the star rating, which would provide flexibility for the outcome domains that have the fewest number of measures. The TEP member also recommended that the outcome measure have two-thirds of the</p>	

TOPIC	DISCUSSION	ACTION ITEM
	weight with the remaining third distributed evenly among the other domains.	

TOPIC	NEXT STEPS	ACTION ITEM
<b>Next Steps</b>	<p>Dr. Venkatesh thanked the group for their participation in the series of TEP meetings and for providing feedback before and after the meetings. He noted that there will be a summary report encompassing each of our TEP meetings and additional questions will be coming to the TEP members in survey form. He requested that TEP members stay engaged and pass along any feedback since the project is still in development.</p> <p>Dr. Venkatesh discussed next steps. He noted that CORE would hold a second public comment period in the spring and a dry run in the summer where additional feedback is expected. CORE will send the TEP the public comment summary report for review when it is made publicly available. He noted that CORE is planning to conduct a focus group type of discussion with coalition members recruited by the National Partnership for Women &amp; Families in order to increase patient engagement. He requested that TEP members consider any questions they may have for patients and consumers and email them to the CMS star ratings inbox. He noted that this inbox would remain active throughout the project.</p> <p>Dr. Venkatesh thanked the group again for all their help and assistance and said that the Star Ratings team looks forward to continuing electronic communication with everybody.</p>	<p><b>TEP members may submit any additional questions or feedback to the CMS Star Ratings inbox (cmsstarratings@yale.edu).</b></p>