Inpatient Quality Reporting (IQR) Program

Support Contractor

Overall Hospital Quality Star Ratings on *Hospital Compare*

Presentation Transcript

Moderator:

Candace Jackson, RN
Project Lead, Hospital IQR Program
Hospital Inpatient Value, Incentives, and Quality Reporting (VIQR)
Outreach and Education Support Contractor (SC)

Speakers:

Pierre Yong, MD, MPH, MS
Acting Director of Quality Measurement and Value-Based Incentives Group (QMVIG) in the Center for Clinical Standards and Quality at the Centers for Medicare & Medicaid Services (CMS)

Arjun Venkatesh, MD, MBA, MHS
Assistant Professor and Director of Quality and Safety Research and Strategy in the Department of Emergency Medicine at the Yale University School of Medicine

Kristie Baus, RN
Technical Advisor, CMS
Technical Lead, *Hospital Compare* Website

May 12, 2016
1 p.m. ET

Candace Jackson: Hello, everyone. Welcome to our webinar. My name is Candace Jackson, and I will be your host for today's event. Before we begin, I would like to make a few announcements. This program is being recorded. A transcript of the presentation along with the Q&As will be posted to the inpatient website www.qualityreportingcenter.com within 10 business days, and it will also be posted to *QualityNet* at a later date. If you registered for the event, a reminder email, as well as a link to the slides, were made available to you about two hours ago. If you did not receive the email, you can download the slides at our inpatient website, again, which is www.qualityreportingcenter.com. And now, I'd like to introduce our guest speakers for today. Pierre Yong is the Acting Director of Quality Measurement and Value-Based Incentives Group in the Center for Clinical

Standards and Quality at the Centers for Medicare & Medicaid Services. Dr. Yong oversees development and operation for a variety of CMS Quality Reporting and Value-Based Purchasing Programs. Previously, he was the director of the Division of Quality Measurement and QMVIG and the Director of Health Care Quality and Outcomes in the Office of the Assistant Secretary of Planning Evaluation at the U.S. Department of Health and Human Services where he oversaw projects on Value-Based Purchasing, quality, comparative effectiveness research, and data infrastructure for patient-centered and outcomes research. Pierre also previously worked at the Institute of Medicine where he focused on a portfolio of projects related to value and cost in healthcare. Dr. Venkatesh is an Assistant Professor and Director of Quality and Safety Research and Strategy in the Department of Emergency Medicine at the Yale University School of Medicine. He is also a scientist at the Yale New Haven Hospital Center for Outcomes Research and Evaluation where he leads the Overall Hospital Quality Star Ratings methodology development team. Dr. Venkatesh's work has focused on the development and performance measures. He's trying to improve emergency department, hospital, and health system outcomes. He has been funded by the Agency for Healthcare Research Quality, the NIH, and the Emergency Medicine Foundation to study care transition and hospital utilization of observation services. He also co-leads the American College of Emergency Physicians, Emergency Quality Network, as part of the CMMI transforming clinical practice initiative. Dr. Venkatesh holds a variety of national leadership roles with American College of Emergency Physicians, the Society for Academic Emergency Medicine, and serves on several Technical Expert Panels for the NQS and AHRQ. Kristie Baus is a registered nurse informatics specialist who works as the *Hospital Compare* team lead. Along with working on *Hospital Compare*, she also functions as a technical adviser working with the various staff to develop and maintain quality measures with several CMS program.

This slide just lists several of the acronyms that will be used in today's presentation.

Today's presentation will educate participants about the methodology used to generate the summary Five-Star rating for individual hospitals using existing measures on *Hospital Compare*, as well as changes to the July 2016 Preview Reports and SAS Pack Distribution.

At the end of today's presentation, participants will be able to describe the overall star ratings methodology, recognize changes made to the July 2016 Preview Report, and access the overall star ratings SAS Pack and supporting materials.

At the end of today's presentation, if time allows, there will be a Q&A session. Any questions that are not answered during the question and answer session at the end of this webinar will be posted to the Quality Reporting Center website within 10 business days. And now, Pierre will begin our webinar. Pierre, the floor is yours.

Pierre Yong: Thanks, Candace. So, welcome to everybody for joining this overview of the Overall Hospital Quality Star Ratings. We are excited to provide this opportunity for you to learn more about the methodology and to answer any questions that you have. If you forward to the next slide.

Today's agenda will cover the overview of the star ratings project, changes to the July 2016 Preview Report, the SAS Pack Distribution (which has been requested), frequently asked questions, and finally close with a question and answer session. Next slide please.

The purpose of the Overall Hospital Quality Star Rating system is to provide consumers with information about multiple dimensions of quality in a single score. So, for those of you who are familiar with *Hospital Compare* know that we display information about individual measures, over 60 of them on *Hospital Compare* for any single hospital. In order to make that information more understandable, we wanted to undertake a

project to develop a single summary rating that was more easily understandable for both patients and their family providers. So, we undertook this project to develop a rigorous scientific methodology for generating a summary Five-Star rating for each hospital, using the existing measures on *Hospital Compare*. Next slide please.

In terms of background behind the development of this project, the current information on *Hospital Compare* can be fairly technical and can be intimidating to beneficiaries and patients. Star ratings, as a concept, have been very widely used in a variety of settings outside healthcare and outside of *Hospital Compare*. It's an easily recognizable and understood way to quickly gain an overall assessment of a particular topic. Patients and consumers have reacted favorably to CMS star ratings efforts, and we have rolled out Five-Star ratings on a variety of our compare sites. And, the Affordable Care Act provided a call for more user-friendly quality information. So, all of these together was driving our desire to move forward on those projects. Next slide please.

We knew that getting stakeholder input on the methodology, as we developed it, was critical to make sure that we did not miss anything, that we realized that we didn't want any unintended consequences, that we shared the work as it's developed. And so, as part of that, we first convened a Technical Expert Panel, which met three times over the course of developing this initial methodology. And, we also had several public comment periods, as well as kept a question and answer inbox, for folks in the public to submit any questions and to provide feedback with us on the methodology that was developed. We also had several discussions with the National Partnership for Women and Families. And, we also did what's called a dry run. So, we've shared with hospitals reports where we ran the methodology against their quality measured data and shared that information with them, as part of their *Hospital Compare* Preview Report, as well as part of this dry run, in order to share with them in advance the methodology and their results based on older data. So, next slide please.

With that, I'm going to turn this over to Dr. Arjun Venkatesh. He's going to give you more details about the project and the methodology.

Arjun Venkatesh: Thank you, Pierre. And, I wanted to thank everyone who's joining us on this call today for the opportunity to talk about the methodology for the Overall Hospital Quality Star Ratings. We've – many of you may have joined us for the call that we held during the dry run last July. We got many great questions after that and subsequently in our Q&A inbox as well. What I'd hope to do today was to review the methodology again, including steps from the beginning and the guiding principles through any changes that were made subsequent to the dry run based on input we received from stakeholders like yourself.

On slide 15, you can see the guiding principles for the development of star ratings. We set these principles at the beginning to ensure that key policy decisions that were made during the course of the development, as well as technical decisions made during the development were consistent with the primary objective of star ratings, which was to summarize overall hospital quality into a single star rating and convey information that's already available on *Hospital Compare* in a straightforward and accessible manner for patients and consumers. To do that, one of the early things we recognized is that we needed to be inclusive of measures. And, we wanted to, therefore, reflect quality at as many hospitals as possible by including as many of the measures on *Hospital Compare* as possible. We also sought to be scientifically rigorous by using established methods, as well as using advance methods statistically that would allow us to summarize these scores. Recognizing that stakeholder engagement was important from the beginning, we've used multiple channels to engage stakeholders both at the project inception, but also as we've become closer to the actual implementation and launch of the star ratings. Part of this engagement with stakeholders has been to ensure transparency through events like this, public comment periods, and a variety of other efforts to engage both technical experts, as well as patients and consumers. And finally, the

overall hospital star rating has sought to develop a lot of consistency with other hospitals – with other Compare sites, as well as with – in terms of their display, but also in terms of the policy decisions that are underlying the star ratings. We've tried to be as consistent as possible in the selection of measures for the Overall Hospital Quality Star Ratings with the weighting of measures that go in to the star ratings and with the groupings of these measures with existing CMS Programs. This also allows the star ratings to evolve and grow over time, as each of those CMS Programs evolves and grows over time.

On slide 16 are a couple of the key considerations we had to think about as we began the development of the methodology. The first was that we were limited to the existing measures on *Hospital Compare*. And so, star ratings, as we look at them, really reflect hospital quality on the existing measures on *Hospital Compare* that are reported by that hospital. What this means is that every quarter, as new measures are added or removed from the *Hospital Compare* website, they're also going to be added or removed from star ratings. Similarly, when specifications to measures are changed and they're updated, those changes immediately get reflected in the star ratings as well. What this means is that when patients or consumers go to the website and see star rating, they're seeing a summary rating of the exact same individual measures that they also see on the website at the same time. And hopefully, this ensures consistency for patients and consumers, as well as reducing any additional confusion. It also means, however, that existing measures may not capture all aspects of quality. While they cover many important and salient aspects of quality, such as mortality or readmissions as well as many important care processes and patient experience, there are likely many other aspects of hospital quality that will be added to *Hospital Compare* in the future and subsequently be part of the Overall Hospital Quality Star Ratings. Another key consideration is, given that we were using the existing or current measures on *Hospital Compare*, we recognize that because of current public reporting requirements, there's a real diversity in the

number and types of measures reported by different hospitals. And so, we sought to develop a methodology that could account for this heterogeneity that some hospitals would report 20 measures, while another hospital would report 40 measures. Or similarly, because of hospitals specializing in certain kinds of care versus other kinds of care, they may be more likely to report cardiovascular measures, say, rather than measures around stroke care.

On slide 17, you can see the general process of the methodology. It's a five-step process that first starts with the selection of measures. Based on all the measures on *Hospital Compare*, we apply the predefined and pre-vetted measure inclusion and exclusion criteria to select 62 of the measures that get included in hospital quality star ratings. These measures have been grouped into seven groups, groups that are parallel to other CMS Programs and to the measure groupings on the *Hospital Compare* website. And then, a statistical model is developed and applied to each of the seven groups. Each of the statistical models seeks to use a Latent Variable approach to calculate a group score or an overall mortality score or an overall safety score and so forth. In step four, those seven group scores are combined into a hospital summary score using a weighted average. Those weighted averages are policy defined and have also been vetted through multiple previous public comments. And then, finally, in step five, a clustering algorithm is used to categorize hospital summary scores into five clusters, each of which corresponds to a star rating going from one to five in whole numbers.

Slide 18 describes several of the changes of the methodology since the dry run. First and probably most importantly, there's been a lot of changes to the measures on the *Hospital Compare* website. Since April 2016, new outcome measures were added around 30-day mortality and readmissions for CAGB surgery. Also, many measures have been retired since April 2016. Two of the retired measures and 14 voluntary measures have been removed from star ratings. Many of these measures run the effectiveness

of care and timeliness of care groups. Also effective July 2016, two measures, OP-29 and OP-30, which are focused on the effectiveness of colonoscopy care will be added to the star ratings. A second change that's happened in the past year has been that there is national improvement on performance of several of the measures. Much of the improvements in readmission rates have gotten a lot of national attention and that is reflective in the star ratings. As hospitals that have improved more so than the national trend have likely seen upward changes in their star rating while those that have improved less so than the national trend have likely seen a reduction in their star ratings. And finally, one technical step was added between the dry run last year and the final release of star ratings and that was the addition of a winsorization step. This winsorization step is added to the hospital summary score prior to clustering, between steps four and five on the previous diagram I showed you. It's done at the most extreme levels, so that only 40 hospitals nationally are winsorized to an extreme value. And, this was based on stakeholder feedback we received during the public comment period in which not only this idea was submitted but it was also very important to patient and consumer groups who sought a broader distribution of hospital quality star ratings.

I'll now walk through each step. On slide 19, you see step one, the selection of measures. The purpose in this step was to be, again, reminding everybody of the original principles of the star ratings methodology, to be as inclusive as possible and include as many measures as possible on *Hospital Compare*. In each quarter, there's a different number of measures that are ultimately available in the *Hospital Compare* downloadable data sets.

In the current quarter, July 2016, that you'll be soon receiving Preview Reports around, there's 115 measures that are possibly included in the *Hospital Compare* downloadable data set. Of those, 13 are planned to be retired or have been suspended or maybe delayed from public reporting. We removed those. There's then three measures that are only reported by

less than 100 of the over 4,500 hospitals on *Hospital Compare*. Because these don't reflect broad reporting of quality on an important measure, we've removed those three measures as well. Third, there were nine structural measures. These are often measures of whether or not a hospital participates in a registry or measures of surgical volume. Based on vetting with our Technical Expert Panel, as well as our first public comment period, because the linkage of these structural measures to outcomes is quite variable, these measures have been removed as well. Six of the measures on *Hospital Compare* are considered non-directional, meaning that it's not exactly clear whether the highest score is the best or the lowest score is the best. This applies to an imaging efficiency measure around mammogram follow-up rates, as well as to several of the spending related measures. As a result, these have also been removed from Overall Hospital Quality Star Ratings. There were 14 measures that are no longer required for IQR or OQR. These are often referred as the voluntary measures. Because they're voluntary, they're not equally reported across all hospitals. To avoid any risk of a response by it, these have also been removed from the Overall Hospital Quality Star Ratings. And finally, there are six measures that are considered duplicates. There are often times, when a measure, such as Influenza Immunizations for Healthcare Workers, may exist in both the IQR and the OQR Program. Or, there may be two measures that are very related, such as the measurement of Hospital Acquired Infections in just the ICU or the entire hospital. When such overlapping measures exist, the primary measure that is broadest and most inclusive consistent with the original principles is used and the duplicate measure is removed. Ultimately, this results in 64 measures being included for the July 2016 star ratings preview period.

Slide 21 describes the processing of these measures. After we narrowed down and select as many measures as possible for our Overall Hospital Quality Star Ratings, we recognized that they are in a variety of different scoring formats, some of these measure time, others reported as percentages, some of these are rates. And so, in order to be able to

combine these measures into a star rating, they need to be standardized. We do this by calculating the Z-score, or the difference between an individual hospital score and the overall mean score for hospitals, divided by the standard deviation. What this ultimately results in is a distribution for each measure that is reported as the actual Z-score. It also ensures that we're able to flip every measure, for which a lower score may be better to a higher score is better, format. So that, for every individual measure score seen in your Hospital Quality Star Ratings Preview Report, a higher score is better. Our second step that we undertake at this point is winsorization of extreme values of individual measures. We set winsorization to set extreme values to the 99.875 percentile or the 0.125 percentile, also known as winsorization to negative 3 or positive 3. What this means is that, for those very few hospitals that may have an individual measure score above that percentile, they get set at the maximum value; or, for those with a few extreme values below that percentile, they get set to that minimum value. In our subsequent analysis, we found that these two steps have no material impact on hospital measurement, but they largely make the modeling approaches later on and the communication of these findings much easier.

Slide [22] begins to describe step two, the grouping of measures. Now that we've included all the measures based on criteria, they've all been standardized and winsorized. They are grouped into seven groups. Three of these groups are outcome groups focused on mortality, safety and readmissions. One group is focused on patient experience. Two groups are of process measures, one on the effectiveness of care and the adherence to evidence-based guidelines, one is on the timeliness of care. And then finally, measures of imaging efficiency of each hospital are grouped together.

On slide 23, we describe this grouping. We recognized that hospital quality is not one monolithic structure and rather that it's represented by several different dimensions; we heard this from our TEP, it's consistent

with prior literature. And so, as we sought to develop groups, we wanted them to be consistent with existing policy groupings and also clinically reasonable. The seven measure groups that we've come upon for Overall Hospital Quality Star Ratings are very consistent with the Hospital Value-Based Purchasing Program. They're consistent with the categories that are used on the *Hospital Compare* website. And, they're also consistent with many other national quality initiatives, including the CMS Quality Strategy. These proposed groups also allow for us to add or remove measures each subsequent quarter as new measures are added or removed to *Hospital Compare*.

On slide 24, you see the number of measures by each group. The largest group is the effectiveness of care group that includes 18 process measures. The smallest group is the imaging efficiency group that includes five measures. And, as you can see, the other groups include between seven and eleven measures. These are database on the July 2016 *Hospital Compare* data set.

On slide 25, you can see the example of a grouping. This is the mortality measure group for July 2016. There are a total of seven measures in this group. It includes measures that have been on *Hospital Compare* for some considerable time, including the 30-day risk standardized mortality rate for acute myocardial infarction, for heart failure, and for pneumonia, as well as newer measures, such as the 30-day risk standardized mortality rate for patients receiving CABG surgery. We also include the death among surgical patients with serious complications in this measure, as the primary intent of that measure was to measure survival.

As another example of an outcome group, on slide 26 you see the safety of care measures group. This includes six measures of Hospital Acquired Infections that are collected through the National Health Safety Network. It also includes a complication measure following hip and knee surgery and the PSI-90-Safety composite measure that is developed by AHRQ.

On slide 27, are the measures that are included in the readmission group. Many of these measures parallel the measures included in the mortality group. And, there are a total of eight measures that are, again, condition-specific, as well as the hospital-wide readmission measure included in this group.

Slide 28 describes the patient experience measures included in the Overall Hospital Quality Star Ratings. As there's already HCAHPS star rating on *Hospital Compare*, we sought to be as consistent with that and include the same individual measure scores in the development of the Overall Hospital Quality Star Ratings. So, for those of you that are familiar with your previous HCAHPS star ratings reports, or have seen the scores before, you'll recognize that each of these measures has what's ultimately called a linear mean score that is adjusted for patient mix and a few other factors. That's the exact same score that's used in the Overall Hospital Quality Star Ratings.

Slide 29 shows measures in the effectiveness of care group. This includes measures of not only Healthcare Personnel Influenza Vaccination, as well as measures of emergency department care. Newer measures have been introduced around colonoscopy care.

And then, many process measures that have been on *Hospital Compare* for a substantial period of time, including measures of stroke, care process, venous thromboembolism prophylaxis.

On slide 31, you see the timeliness of care measures. These include several measures of ED throughput, as well as timeliness in certain care process, such obtaining an EKG in patients with chest pain or managing pain in patients with long bone fractures.

On slide 32 is the final group. These are measures from the hospital OQR program focused on imaging efficiency, covering a wide range of indications.

Step three is largely the most technical component of the Overall Hospital Quality Star Ratings, and it's seen on slide 33. In step one, the measures were selected. In step two, we group them into seven groups. And, in step three, we use statistical models to calculate a group score for each of these groups.

Slide 34 describes this methodology. We apply a Latent Variable Models as an analytic approach that seeks to measure each of these dimensions of quality. The resting assumption is that each of these aspects of quality cannot be directly measured, but that we can learn something about each of these aspects of quality based on the existing measures on *Hospital Compare*. Another way to say that is to say that in the case of mortality, a hospital is likely to have a variety of aspects to care that allow for reduced mortality or improve survival for patients. We know something about that mortality from the condition-specific mortality measures that are contained in each one. There's common information that goes across those measures. And, that common information reflect that Latent or unmeasured aspect of quality. And so, what we seek to do is use a statistical model to estimate or calculate that Latent Variable or that Latent aspect of quality that was otherwise unmeasured. Each of these statistical models are separately created for each measure group, and they're designed in a way to accommodate several differences in the *Hospital Compare* data. They're setup to account for missing information and recognizing the fact that some hospitals may have three or four mortality measures, while another hospital may have seven mortality measures. They also accommodate diverse hospital reporting patterns, that the measures may be different, even if the measures are of the same number between different hospitals. They also allow for the addition and removal of measures over time. In our initial testing, we did analyses that showed that each of the groups used within the Overall Hospital Quality Star Ratings, each reflect a different aspect of quality. And subsequently, that allows us to keep specific or similar measures within each group. The final thing that's probably most important to the Latent Variable Model is

that they seek to find consistency and they look at the relationship between measures. Instead of simply doing a weighted average of the measures within a group of even a simple average of measures within a group, the Latent Variable Model seeks to find measures that are more correlated with each other, assuming that those measures tell us a little bit more about that latent aspect of quality. It also removes the subjective assignment of weights to individual measures by instead allowing the data that drive the weighting of individual measures.

On slide 35, we describe the accounting of sample variation in the methodology. For each measure on *Hospital Compare*, in addition to a hospital having an individual measure score, they often may have a different number of cases. One hospital may have 75 patients with heart failure included towards their re-standardized readmission rate, while another hospital may have 35 or 40 patients. The star ratings methodology accounts for this variation. Hospitals that have a larger denominator are determined to have a more precise measure score; and, therefore, are weighted more so in the model. This is a method known as weighted likelihood. And, it allows us to know when we have more information to make a more precise assessment of a hospital's overall star rating in comparison to hospitals that may have smaller denominator; and, therefore, a less precise score.

On slide 36, we described the concept of measure loading. I alluded to this earlier when I mentioned the value of the model in looking for common information between measures. Another way of describing this is saying that the Latent Variable Model estimates a loading for each measure in a group associated with the hospital-specific group score. What is that loading? It's the association between an individual measure and the group score or that dimension of quality relative to all other measures in the group. What that means is that it's calculated for the entire group at once within the model. And so, a measure's loading is the same for every hospital. So, as you look at your reports and you see a

measure loading that is read as .32, it's .32 for every hospital in the country. And, measures that have higher loading are measures that when we model the data had a stronger association with the group score. They're not proportional weights as people often think about weights, and so, you can't just take a loading multiply it by the measure score, add them up and get your group score. Rather, they are produced at the same time that your hospital's group score is calculated and they reflect what we see as the underlying relationships in the data. There's nothing arbitrarily or subjectively set about loadings, they are just the natural outgrowth of the data, as they currently stand.

Slide 37 is where we begin to describe step four, the generation of a summary score. As a reminder of where we're at, in step one, which has over 100 possible measures on *Hospital Compare* and for July of 2016 selected 62 that are going to be included in the hospital quality star ratings. In step two, we grouped those measures into seven measure groups. In step three, we used seven separate statistical models to develop a measure group score for each group. And now, in step four, we used policy-based weighting to combine those measures into a single hospital summary score.

Slide 38 describes this weighted average process. Here you see each measure group listed, as well as the set measure group weight. For the three outcome groups, they're weighted at 22 percent, as is the patient experience group. And then, the two process groups and the imaging efficiency group are weighted at four percent. These weights were initially developed to be consistent with the CMS Hospital Value-Based Purchasing Program and to be consistent with the CMS Quality Strategy that seeks to emphasize outcomes over processes. We took these weights to our Technical Expert Panel for additional feedback. These weights were presented to a patient and patient advocate workgroup, we worked with through the entire measure development process, for feedback. And, they were also vetted through a public comment process. We broadly

received very supportive comments of these weights, and they're ultimately finalized for you in the Overall Hospital Quality Star Ratings methodology.

On slide 39, you can see the criteria by which we developed these weights. It was by thinking about the importance of these measures in various policy programs, as well as various multi-stakeholder frameworks of quality. It was our goal to remain consistent both within *Hospital Compare* but also across the goals of CMS. The third reasoning was around policy priorities, which seek to emphasize outcomes over process measures. And then, as I mentioned, fourth, stakeholder input, was critical to the development of these weights, as well as the vetting of these weights. The weights that are used in the July 2016 Overall Hospital Quality Star Ratings Preview Reports are identical to the ones that were also seen in last year's hospital dry run.

On slide 40, we describe the relationship between these weights and the FY 17 Hospital Value-Based Purchasing weights. As you can see, outcome groups are weighted highly and very similarly. Readmission measures are not part of the Value-Based Purchasing Program; and so, we had to consider how to best incorporate that outcome group into this methodology. Similarly process measures are all grouped into a single measure in the Value-Based Purchasing Program, and we created distinct groupings there of equal weighting for process measures. And finally, because the overall star rating does not currently include measures of cost, the single Medicare Spending Per Beneficiary measure in the Hospital Value-Based Purchasing Program is not included. And, that weight is reallocated across other outcome groups.

Slide 41 describes what we do when a hospital may not have all measure groups. As I mentioned earlier, the average hospital on *Hospital Compare*, they report around 45 measures with a broad range of reporting based on the hospital size and case mix. If a hospital doesn't have any measures in a group, that group is then considered missing. It doesn't

count towards that hospital's overall hospital quality star ratings. Any group, however, that even has one or two measures, will still result in the calculation of a group score. And so, the star ratings methodology uses the same approach as the Hospital Value-Based Purchasing Program when there are missing groups or hospitals that have groups that have no measures. And, by doing that, we redistribute the weights to the non-missing measure groups.

On slide 42 is an example of this redistribution. You first see the standard weights. Again, 22 percent for the outcome groups, four percent for the process groups. In this example, the hospital doesn't have any measures in the efficient use of medical imaging group, therefore, they don't have a full 100 percent to work with. They've gotten that 96 percent to work with. And so, that four percent that remains is redistributed proportionally across the other groups, so that the relative relationship between outcomes and process and between each outcome group maintains the same percentage.

Slide 43 describes the final step, the use of a clustering algorithm to summarize the hospital summary score into five star categories. These are whole star ratings ranging from one to five, consistent with the other CMS Compare sites.

Slide 44 describes this categorization process. First, hospital summary scores are winsorized at the most extreme values. CMS analyzes the distribution of hospital summary scores and winsorizes it to the .5 and 99.5 percentile. This results into the 46 most extreme hospitals summary scores being winsorized to those either minimum or maximum points.

We then apply k-Means Clustering. k-Means Clustering seeks to minimize the sum of squares of distance between a hospital's summary score and each cluster mean score. Another way of saying that is it tries to cluster hospitals so that hospitals have the most similar score grouped in a cluster and least like the adjacent group. As an example, if you thought

about four-star hospitals, these are a cluster of hospitals with higher than the average summary score. Hospitals in that cluster have scores that are most similar to each other, but in a way so that the difference between the hospitals in that group is less than the difference between hospitals in the four-star group and hospitals in the three or five-star group, the adjacent bucket.

Slide 45 demonstrates preliminary data for the July 2016 *Hospital Compare* results. What you see is that like in other star ratings programs, most hospitals in the country are in the three-star category, just below 50 percent. There are many hospitals in the four-star category more so than the two-star category. And then a few hospitals, 100, are in the five-star category nationally.

Slide 46 describes the thresholds for public reporting, recognizing that some hospitals may report very few measures; and, therefore, may have unreliable estimates of their measure group scores or less face validity. A minimum threshold was created to ensure that only hospitals above that line would be included in Overall Hospital Quality Star Ratings. This minimum reporting threshold is similar to what is used for Hospital Value-Based Purchasing, which is based on both reliability and face validity. This method was vetted with our Technical Expert Panel, with our patient and patient advocate workgroup, and was also included in the public comment. We received broad consensus support for the reporting threshold. We set it at, first, a minimum of three of the seven measure groups being reported, of which one must be an outcome group, meaning mortality, readmission or patient safety. Second, within the three included groups, each hospital must also have at least three individual measures in order to be included in the Overall Hospital Quality Star Ratings.

Slide 47 shows data that we showed during the hospital dry run, national stakeholder call that may remind you of how we determined the number of minimum measures and minimum groups required for public reporting. As you require more measures or more groups, you can see a fewer

percentage of hospitals would be included. For example, if we had an extreme threshold where all seven measure groups were required, and four measures were required per each of those groups, only 39 percent of the hospitals in the country would receive a star rating. On the converse, if we set the threshold so low that only two measure groups were required with only one measure per group, then almost every hospital, 97 percent in the country, would receive a star rating. Ultimately, the rule we determined is seen in the cross between the three minimum measure and the three minimum measure group percentage, which during the April 2015 dry run resulted in 3,709 hospitals, or 78 percent of all hospitals, receiving a star rating.

In slide 48, we described another addition that was added to the Overall Hospital Quality Star Ratings based on feedback we received from stakeholders during both the dry run period, as well as from patient and consumer feedback. For measure groups, in which at least three measures have been met, CMS has also reported categorical group performance. These categorical scores are meant to complement the overall hospital quality star rating and complement individual measures. Recognizing that often for certain patients, one aspect of quality may be more important than others, these categorical scores can help provide more detail for patients and consumers. These are calculated by comparing the hospital's group score and the 95 percent confidence interval around that score to the national average. If a hospital's 95 percent confidence interval is entirely above the national average score, then that hospital is given a group category of above the national average indicating better performance on measures within that group. For example, measures of patient safety. Hospitals that have a confidence interval that overlaps with the national average, receive a group score of same as the national average. And then, finally, hospitals with a confidence interval that falls entirely below the national average are given a category score of below the national average. This methodology is consistent with what is done for many individual measures on the *Hospital Compare* website.

Slide 49 describes group performance categories for each of the seven measure groups for July 2016 public reporting. As you can see, for most groups, most hospitals fall as the same as the national average. However, for several of these groups, there may be more hospitals that are skewed towards a higher performance rate or above the national average because of the underlying distribution of measures.

On slide 50 are a few items to note regarding implementation. The star ratings methodology is designed to be refreshed on a quarterly basis in 2016; but then, to be refreshed on a semiannually basis in 2017 and onwards, as a result of modifications being made to the overall *Hospital Compare* website. In addition, the star ratings results are going to be located on several pages on *Hospital Compare*. In addition to being part of overall results pages, these results may be found on *Hospital Compare* pages and Hospital Profile pages. Do note that the group categories, which have been added to *Hospital Compare* in response to stakeholder feedback are going to be located on the view details page, which is a click – simple click, or tab from the overall star rating found on the general information tab.

Slide 51 describes resources for those that may have additional questions. Over the past several months, during the initial implementation over our hospital quality star ratings, we received many great questions from you all through our inbox, and we'd encouraged continued questions both about your individual reports, as well as ideas that may exist for the methodology as it continues to evolve over time. There are also many resources specific to methodology that may not have been covered in this call, on the *QualityNet* page. This can be found under the Hospital Inpatient Quality Reporting Hospital Star Ratings page. The one request we do have, if you ever ask any specific questions or submit us any of your hospital data, is that you please do not include any PHI in these comments or questions. With that, I'll close, I'll thank everyone for walking through the methodology with me. And, I'm going to turn it over

to Kristie Baus to discuss upcoming changes to the Preview Reports and additional considerations around implementation. Thank you.

Kristie Baus: Thank you, Arjun, for the presentation on the methodology behind the star rating. My name is Kristie Baus, I am the *Hospital Compare* Team Lead here at CMS, and I'm going to talk a little bit about the changes you will notice to your July 2016 Preview Reports.

The first thing you'll notice is that the weights and the group scores will be shown for any measure group that is included in the calculation of the summary score. This allows hospitals the ability to recalculate the summary score, the group scores and the weight.

On slide 54, you will notice differences between the two Preview Reports. The top shows the Preview Report for April public reporting. The bottom shows the Preview Report for July public reporting. Again, group scores and weights will be shown for all groups that have applicable measures used in the star ratings.

Along with Preview Report, hospital-specific reports will also be provided starting with the July 2016 *Hospital Compare* release. These hospital-specific reports are similar in nature to what hospitals already see for outcome measures. The HSRs are designed to complement data in the Preview Report and to provide additional details for hospitals, and will include the confidence intervals for summary scores and group scores as well as the individual standardized measure scores. A release date of these hospital-specific reports is to be determined, but they should be released shortly.

Next, I will go over about this SAS Pack Distribution.

On slide 57, we'll talk a little bit about the SAS Pack. In response to the multitude of stakeholder feedbacks and to ensure continued transparency of the star ratings work, CMS is currently considering the feasibility of providing the star ratings SAS Pack in applicable input file to hospitals.

The SAS Pack includes all the relevant SAS code and documentation, including user guide, as well as the input file which include all the hospital data needed to replicate star ratings, as well as the national denominators. We will also release an instruction guide for how to use the SAS Pack. It is important to note that the release of the SAS Pack also requires the release of all hospitals' individual data. At this time, we are currently exploring the feasibility of releasing this data to hospitals. And, we'll keep everybody posted. It's also important to note that using a different version of the SAS software, as well as different hardware, within your computer system may result in different scores.

It is possible that hospitals get different results if the system requirements are not identical to those provided in the SAS instructions. Manipulation of the code or the input file can also result in varying results. Help desk support will not be available at this time to answer technical questions. We encourage hospitals to refer to the instruction guide for system requirements and a step-by-step guide on running the SAS Pack. FAQs will be provided on the Hospital Star Rating *QualityNet* page to provide SAS Pack guidance.

And now, we'll go over some frequently asked questions.

On slide 60, you'll see the most common frequently asked questions that CMS has received, both through stakeholder correspondences and through our star ratings inbox.

The first question is, can the release of the star ratings be delayed until the full impact on hospitals is analyzed? In order to give more time for hospitals and other stakeholders to become more familiar with the methodology, CMS has postponed the initial public reporting of the overall hospital star rating until the future release yet to be determined. In the meantime, CMS will continue to work with stakeholders and hospitals to learn about their data and analyze their questions. We also encourage

hospitals to use the star ratings mailbox to provide feedback and comments on how we can evolve *Hospital Compare* in the future.

And, why is my measure group score for mortality and/or readmissions below the national average, when my individual measures are the same as the national average or above the national average? The methodology ensures all included measures are in the same direction. For example, a higher score indicates better quality. The methodology combines these measure values into a group score. A hospital score may be substantially lower than the national average, but not statistically lower on several measures. When combined, this shows consistently lower performance, resulting in a lower category score.

For each measure group, the 95 percent confidence interval of a hospital's group score is compared to the national average to assign a national comparison category according to the following guidelines. The measure group score does not directly translate into a national performance category, since the 95 percent confidence interval is required to compare the measure groups to the national average. This is the same methodology CMS uses for our outcome measures, as well as hospital associated infection data.

Why is my production star rating different or lower than my dry run star rating? A change in the overall summary score and star rating can be attributed to a number of factors. The measures used in the calculation for the overall summary score and star ratings have changed. The methodology has been updated since the dry run to include winsorization. The hospital's performance may have worsened or the national performance may have improved beyond the hospital's performance. In addition, the minimum and maximum hospital summary scores for each star category will change with each reporting period based on the underlying distribution of hospital summary scores. Next slide please.

Can individual quality measures be removed? CMS believes that the PSI-90 composite is an important measure of overall hospital quality. Several indicators within the composite are not included in other measures. The measure addresses issues that are harmful to patients, with limited burden on hospitals for data collection. CMS developed the star ratings to be as inclusive of as many measures as possible on *Hospital Compare*. Methodology includes systematic process for the inclusion of measures. All measures in the methodology pass the selection criteria.

By studying outlying hospitals, CMS concluded that it is unlikely that any one measure precludes a hospital of a given type from performing well on the star ratings. CMS will continue to consider your feedback in evaluating future measures, including the patient safety indicators.

Are star ratings adjusted for sociodemographic factors? The Overall Hospital Quality Star Ratings represent a performance summary based on individual measures already reported on *Hospital Compare*. CMS is committed to addressing concerns about the role of SDS factors for individual measures, improving outcomes, and working with stakeholders to improve individual quality measures.

The Office of the Assistant Secretary for Planning and Evaluation, or ASPE, is conducting research on risk adjustment for SDS, as directed by the IMPACT Act. ASPE will issue a report to Congress by October 2016. CMS will examine the recommendations issued by ASPE and consider how they apply the publicly reported measures.

Can the overall star ratings be misleading to patients and consumers? The overall star rating represents a performance summary designed to facilitate patient and consumer use of *Hospital Compare*. It responds to sections of the Affordable Care Act, which calls for public reporting that is transparent, efficient, easily understood, and widely available. The data used in the overall star rating is the data available on *Hospital Compare*. CMS consulted both with Technical Expert Panels and patient advocate

working groups, which includes diverse patient and patient advocate representation. A patient advocate group supports CMS' decision to provide a hospital quality star rating system.

CMS welcomes stakeholder feedback and appreciates your continued and thoughtful engagement to CMS' Quality Measurement Work.

And now, I will turn it over to Debra Price, who will walk us through the question and answer session.

Debra Price: Well, thank you for that introduction. And now, I will start talking about the continuing education credits. This is Debra Price. Today's webinar has been approved for a 1.5 continuing education credits by the boards listed on the slide. We are now a nationally accredited nursing provider. And as such, all nurses report their own credits to their respective boards using our National Provider Number shown on the last bullet here. It's number 16578. It is your responsibility to submit this form to your accrediting body.

We now have an online CE certificate process. You can receive the CE certificate two different ways or two different times. One, if you registered for the webinar through ReadyTalk®, you will get a survey at the end of our slides. The survey will allow you to get your certificate. However, you will only be able to get that certificate if you were the one that registered. The second way to get a certificate is within 48 hours, we will be sending out a separate survey. When you receive the survey, please give people who are in your room listening but did not register through ReadyTalk®, please give them the survey. They take the survey and then they will get the certificate themselves. After the completion of the survey, you click the "Done" button at the bottom of the page and another page will open. You will need to choose to register as either a new user or an existing user. If you've been receiving certificates with us all along, and you haven't had any problems, go ahead and click on the "existing user" link. If you have never received a certificate, or if you had

problems in the past getting your certificate, please register as a new user using a personal email. Just to note that healthcare facilities have firewalls that are continuously being upgraded. And, you may have a firewall up on this even that wasn't up last week, if you've attended any of our other events.

If you do not immediately receive an email to the address you registered with after the survey, that means that there is a firewall up. And, what you'll need to do is go back and register as a new user, using your personal email address.

This is what the survey will look like. It will pop up again at the end of the event. And, again, we will send you a survey within 48 hours. You see, on the bottom right hand corner, the little "Done" button, that's what you're going to click on when you are finished with the survey.

This is the page that pops up when you click the "Done" button. This is what I was talking about previously, where you have two links, a new user link and an existing user link. New user is, if you have never gotten a certificate from us or if you've had problems in the past getting a certificate. Use the new user link and make sure you fill in the form for your personal email. If you have been receiving certificates all along, please click on the "Existing User" link.

This is what the new user screenshot looks like. So, if you click on the "New User" link, you put your first name, your last name, your personal email, and a phone number that will be identified with the email.

Remember, again, to use a personal email because hospitals and other healthcare facilities have firewalls that are constantly changing and being upgraded.

This is what the existing user screen looks like. If you've been receiving certificates all along, please fill in your username, which is your email address complete with whatever is after the @ sign. So, it will be your

complete email address and whatever password you used when you registered. And, if you don't remember what your password is, then you'll have to get back with us, and we'll have to reset your password.

And now, I will pass the webinar back to your host. I hope you do not have any problems getting your certificate. If you do, my email will be on the survey today, as well as the survey you're going to receive in 48 hours. Thank you for your time and have a great rest of the day.

Candace Jackson: Thank you, Deb. As Deb indicated, we will now go into our question and answer session. We have had a lot of questions submitted through the chat feature, and we will get to as many of them as we can. So, if your question does not get responded to today, please remember that all questions will get a response provided and will be posted to the Quality Reporting Center website at a later date. So, we will go ahead now and start with the questions and answers as time allows. And, our first question is: I would like to know how CMS is going to address the issue of the lack of social demographic adjustments for the readmissions and other outcome measures? This is particularly important for rural community hospitals.

Kristie Baus: Hey, Candace, thank you. This is Kristie Baus from CMS. And, to answer that question, which was also covered in the frequently asked questions: CMS is currently working with other federal partners, including ASPE, the Assistant Secretary for Planning and Evaluation, to determine the steps to take to adjust for the socioeconomic factors. ASPE is due to be – to release a report in the fall of this year. Once that report is released, CMS will look at that report and determine if or when our measures would be adjusted for socioeconomic factors. After that, then it will be reflected ultimately in the star ratings.

Candace Jackson: Thank you, Kristie. And, just to our panel who is going to be answering the questions and comments: some of these are very long, so if you do need me to repeat them so that you have a better understanding of the

question or comments, please let me know, and I will repeat the question. Our next question: as a hospital system, composed of both medical center and rural setting hospitals, we are concerned that medical centers are at a disadvantage because of the increased number of patients, more measures included and sicker patients who are more at risk for some of these complications. Likewise, rural setting entities are disadvantage, as we know zip code accounts for large percentage a variation in hospital readmissions. Are you able to respond to that comment?

Arjun Venkatesh: Hi, this is Arjun Venkatesh here from Yale. Thank you for this question. I should admit that I'm not entirely well versed in what these proper noun medical centers, but I think that in terms of how hospitals – different types of hospitals – or hospital locations may interact with star ratings, I think the two things that – from the methodology perspective, that probably apply here are that we designed the methodology really agnostic to type of hospital at the beginning. And in doing so, we did that because conceptually what we wanted in the methodology was something that would reflect the information already on *Hospital Compare* in each individual measure and to not, you know, further defined or further change any of that information and cause any additional confusion. We have done some preliminary analysis earlier on, looking at several hospital characteristics or types of factors and how that may be related to star ratings. And, I found a broad range of star ratings across many hospital characteristics including (urbanicity or rurality) as well. And this is an area I know that CMS is continuing to explore and will look to their guidance as we do further analysis.

Candace Jackson: Thank you, Arjun. Our next question: are hospital's overall hospital star rating outcome readmissions group score is below the national average when the 30-day risk standardized condition-specific and hospital wide readmission rates are no different than the national rate. Why is this so?

Kristie Baus: So, this is…

Arjun Venkatesh: So…

Kristie Baus: …Kristie – oh, I'm sorry. Go ahead, Arjun.

Arjun Venkatesh: So, this is a question we've also received in our inbox over the past few months, associated with Preview Reports that people have seen. And, probably the easiest way to think about this is to remember that individual measures, when they are reported, will have a different national average score and a different 95 percent confidence interval that needs to be exceeded, or be below ~~of ,in~~of, in order to be called above or below the national average. In the case of a group score, what we have to remember is that we start to combine and summarized information across many measures. And so, if a hospital does well, and let's say moderately well at one readmission measure, it may be possible that it didn't meet the threshold to be above the national average for that individual measure. But, if they do moderately well across all of the readmission measures, and some in combination, that might be very high performance, in comparison to other hospitals; and, therefore they may end up with a group score that's above the national average. And so, this is one of the features of the star ratings performance categories that seeks to show and summarize that performance when there's kind of broad performance that is strong or poor on either end of the spectrum.

Candace Jackson: Thank you, Arjun. Kristie, did you have anything else to include in that response.

Kristie Baus: No, I think Arjun covered it, thank you.

Candace Jackson: OK, thank you. Our next question: since the mortality metric has been standardized and now has a reverse direction, should the name be changed to survival rather than mortality? My institution is above average mortality, which on the surface sounds bad. But, if we had above average survival, that would sound good.

Kristie Baus:	OK, this is Kristie. Thank you for that suggestion. It is something that the workgroup has discussed, and we will definitely take that under consideration.
Candace Jackson:	Thank you, Kristie. Our next question: will CMS share the calculation workbook, so we can supply our most recent data into the workbook to anticipate what the star rating will be? The star rating CMS publishes usually uses older data.
Kristie Baus:	I'm not 100 percent sure what workbook you're referring to. I will say that currently CMS is exploring the best way to release the statistical software package and the code used to calculate the star ratings. That is – you know, again, that's something that we are exploring the feasibility of doing. And, when we do, we will definitely send out notifications to all hospitals so that they can run their own data.
Candace Jackson:	And, this next question may piggyback on to that, it says: is it possible for hospitals to replicate the star rating calculations in order to better monitor our performance ongoing?
Kristie Baus:	Right. That does sound like the same question. And, again, we are exploring the feasibility of releasing that statistical software packaging code and the national input file.
Candace Jackson:	Thank you, Kristie. Our next question: do all the *Hospital Compare* measures use Top-Box scores?
Arjun Venkatesh:	So I'm…
Kristie Baus:	I…
Arjun Venkatesh:	Go ahead, Kristie.
Kristie Baus:	I think the only measures that used the Top-Box scores are the HCAHPS measures.

Arjun Venkatesh: Yes. And, in the case of the star ratings methodology for the HCAHPS measures, the scores that are used or referred to as linear mean scores. And so, those are actually not the Top-Box scores, but they are continuous numbers, and they're the same scores that are used in the HCAHPS star rating. I believe – I don't know the details fully of these measures, but they are adjusted for things such as for certain patient level factors.

Candace Jackson: Thank you. Our next question: we have a question on PSI-90-Safety measures. AHRQ submitted a revised measure set for the PSI-90 for approvals to NQF. In its submission, AHRQ identified several significant biases and flaws in the current methodology. This information became available after the advisory panel meetings occurred. Can you explain why CMS has not eliminated this variable from the star rating, methodology given the known limitations and impact on the star rating?

Kristie Baus: Hi, this is Kristie. We did receive a lot of correspondence around the PSI-90 measure. And, at this time, we are working with our star ratings team to do some analyses on the measure and determine the best way to move forward. The hospital star ratings were designed to be as inclusive of as many measures that are currently publicly reported on the *Hospital Compare* website. And so, that's why, you know – that's, you know, one reason why the PSI-90 measure has not been removed from the star ratings. We are – we are in tuned with what AHRQ is doing. The measure remains NQF as endorsed. And at this time, we do not have plans to remove PSI-90 from the star rating calculation. But, again, we are doing further analysis.

Candace Jackson: Thank you, Kristie. Our next question: how often will measures be added/removed? How are these being shared with hospitals and other key stakeholders?

Kristie Baus: So, one of the slides that I believe Arjun went over had to do with the inclusion and exclusion criteria for the star ratings. When measures are added to *Hospital Compare*, they will be included in the star rating

calculation, if they meet all the criteria; that is, if they're not like a structural measure. It's any measure that's a measure of quality will be included in the star ratings. If a measure is retired or removed from the inpatient or outpatient quality reporting programs, they will then be removed from the star rating calculation as well. The star rating – I think it's important to know, the star rating is meant to be a summary depiction of the measure data on *Hospital Compare*.

Candace Jackson: Thank you, Kristie, for your response. The next question: with the latest variable models, what evidence can be provided about the step to which a single Latent Dimensions accounts for common variance amongst targeted measures for each LVM?

Arjun Venkatesh: Thanks for this question. Early in the development process, when we sought to evaluate whether or not the Latent Variable Modeling approach would both be appropriate, as well as kind of meet the objectives of the star ratings, we did several factor analyses. And, what we found in our factor analyses was that the use of these measure groupings, as they're currently used, identifies one meaningful Latent Variable per each group. And so, I think that the assumption that the mortality measures together, for example, all reflect one common Latent Variable with respect to mortality performance was fairly strong and robust. The one exception to this was the measures in the efficiency group, where there may have been one to two predominant Latent Variables. That group is weighted very little towards the overall star rating. And, when we vetted it with both our multi-stakeholder Technical Expert Panel, as well as the public through public comment period, the general consensus was that the principle of inclusiveness of measures meant that we should still include those measures and include that group as a distinct group.

Candace Jackson: Thank you, Arjun. Our next question: is the Five-Star rating limited to PPS hospitals or are Critical Access Hospitals included?

Kristie Baus: Any hospital that has data on *Hospital Compare* and meets all of the inclusion criteria for measures, you know, has at least one of those outcomes categories, can receive a star rating.

Candace Jackson: Thank you, Kristie. Our next question: will CMS provide a tool calculator to aid hospitals in predicting or modeling their star rating on a more concurrent process?

Kristie Baus: Again, we are exploring the feasibility of releasing that statistical software for that purpose.

Candace Jackson: And, the next question: what is the timeframe for the July report that will be reflected in the stars?

Kristie Baus: You mean – I wonder – I think they're probably referring to the measure dates, that would be my guess in this question. So, the measure data for the process of care measures, the measure dates will be one quarter more – or advance one quarter, from where they are now. So, for example, it would be, let's see, third quarter '14 to second quarter – I'm sorry – yes, second quarter of '15 data. So, the process measures HCAHPS and HAIs always roll forward a quarter. For the outcome of care measures in July, the majority of them – these are the measures that use three years of data, they will encompass – and I'm looking it up here – July 1, 2013, to June 30, 2015.

Candace Jackson: Thank you, Kristie. Our next question: k-Means Clustering is typically used when classifying objects based on multiple measures. In this case, k-Means was applied to a single summary score. How does this effectively differ from breaking the summary score distribution into five ranked categories?

Arjun Venkatesh: Thanks. So, the k-Means Clustering approach was one that we considered in the methodology development process, alongside a few other options. I acknowledge that often k-Means is used to classify or cluster variables, when you've got multiple variables, but it still can work with a single

dimension variable, or one variable, in this case the hospital summary score. So, to simplify this a little, I guess I should say that when we originally developed the methodology, we considered several ways to classify or cluster hospitals into each of the star categories. The most simple way you could imagine that we thought of was to set five lines, or essentially making it into quintiles. Every hospital from the zero to the 19.999 would be the first quintile, and that would be one star. From 20 to 39.999 that would the second quintile, and that would be two stars. Another approach we considered was to set statistical thresholds and say that a hospital had to be – have a hospital summary score that was statistically higher than the national average score, and that that also had to be true for a certain number of measures and number of groups contained within the summary score. The third approach we considered, which is what we ultimately used is the k-Means Clustering approach. We took all of these approaches to the Technical Expert Panel, as well as a discussion and public comment. And, the general consensus was that the k-Means Clustering would allow us to meet a variety of goals in that classification that the others didn't meet. For one thing, it didn't create an arbitrary – as arbitrary of the line between something, say, like nineteenth and twentieth percentile. Also, those hospitals may have a score that is so very similar and near identical, but they'd be getting a different star ratings simply because the line was made at a fixed equal point. k-Means Clustering allows for the size of each group to be unequal. It allows for there to be, in the case of this most recent reporting period, more three star hospitals because many hospitals perform near the national average overall across all the measures that they report. The other advantage of k-Means Clustering is that it intuitively groups hospitals together that have a hospital summary score that is more similar. And so, when we've done subsequent testing, testing things such as the validity or using simulation to test the reliability of our classification of hospitals, we find that it would perform better than something different, such as the quintiles approach. I think this is a place where CMS has also said, in the past, that they're open to additional feedback and comments. It's likely a part of the

methodology that will continue to evolve over time. But, we think, at this initial point, it was a good place to start with, in terms of classifying hospitals into these five groups.

Candace Jackson: Thank you, Arjun. Our next question: any thoughts on why all Five-Star hospitals in the HCAHPS summary are under 400 beds predominantly for profit and suburban? Can we expect a similar distribution in the overall star ratings?

Kristie Baus: We are currently doing some analysis on the distribution of the star ratings, and those results will be forthcoming at a later date. I can't answer that question at this time.

Candace Jackson: Thank you, Kristie. And our next question: where do we download the Preview Reports on *QualityNet*? Are these for IQR only or combined with OQR Preview Reports?

Kristie Baus: So, the Preview Reports – I don't have the instructions on how to download Preview Reports from *QualityNet*. But, you can contact the *QualityNet* helpdesk, and I'm not sure if you might have the email address there, Candace, that you can share, and they can walk you through how to download your report.

Candace Jackson: We can provide that email address along with - or that address along with the responses.

Kristie Baus: That would be great, thank you, Candace. And just – and so everybody is aware, the reports will be available through June 6 – I believe it's June 6. And, we do encourage hospitals to download your reports and review them.

Candace Jackson: Thank you. And, our next question: will the performance on the four eCQMS required for 2016 IQR be included in future star ratings?

Kristie Baus: That's possible. Again, we're – the star ratings are designed to be as inclusive of as many measures posted on the *Hospital Compare* website. So, at a time when CMS does decide to add the results of the eCQMs, we'll determine then whether or not they'll be included in the star ratings.

Candace Jackson: And our next question: what would be the best method or process to take to determine why a star rating may have changed?

Kristie Baus: I think, you know, just looking and comparing your preview reports from the previous release and the current release. And then, taking into consideration performance on any of the new measures that may have been added, as well as considerations of performance of any measures that may have been removed from the programs, and then ultimately from the website.

Candace Jackson: OK, thank you. And, our next question: are these rates risk-adjusted rates, such as the heart failure or mortality rates?

Kristie Baus: The star rating in and of itself is not risk-adjusted. If the underlying measures are risk-adjusted, then, you know, we didn't do anything to change the calculation of the underlying measures. And, I don't know, Arjun, if you have anything else to add to that or…

Arjun Venkatesh: No, I think that's a good summary. I think that the star ratings have been designed, as we mentioned earlier, to not – to just summarize the available information. And so, introduction of risk adjustment into star ratings, but not individual measures, would potentially either further complicate things, the technical properties would be unclear. So, what we've designed it to do is really be something where risk adjustment, when it – whatever the factor may be, is really – should be designed at the individual measure level, for that individual measure's validity and reliability.

Candace Jackson: Thank you. And, we have time for one more question. How did you determine the rating of the measures? For instance, it may not be

intuitive, or reasonable, that an outcome, such as mortality, should carry equal weight to readmissions.

Arjun Venkatesh: So the weight – the methodology development process around the weighting involved multiple, I would say, forms of feedback. When we initially developed the weights, we tried to use places where there was already some policy guidance. And so, there are weights that are used in the Hospital Value-Based Purchasing Program that emphasized outcomes over process and equally emphasize different domains of outcomes. We took those weights, and we vetted with them the multi-stakeholder Technical Expert Panel, we showed it to a patient and patient advocate workgroup, and there was also public comment on them in the spring of last year. We received broadly, I would say, a lot of support for these weights, in which outcomes are emphasized over process, and each of the outcomes groups were equally weighted. There's no gold standard or no correct or right number for weights, but these seems to be consistent with a variety of policy programs and a place where there was initial consensus. And, it's another place where, if folks have additional feedback or additional concerns, that's something, that from a methodology development perspective, we're certainly very open to.

Candace Jackson: Thank you, Arjun. And that concludes our webinar for today. I'd like to thank Pierre, Arjun and Kristie again for being speakers on today's webinar and providing this useful information. We hope that it has proven very beneficial for you. We hope that you have a great day and thank you very much for attending.

END