MLN Connects®

National Provider Call Transcript

Centers for Medicare & Medicaid Services
Hospital Compare Overall Star Ratings Methodology
MLN Connects National Provider Call
Moderator: Aryeh Langer
August 13, 2015
1:30 p.m. ET

Contents

Operator: At this time, I would to welcome everyone to today's call. All lines will remain in a listen-only mode until the question-and-answer session. This call is being recorded and transcribed. If anyone has any objections, you may disconnect at this time.

I will now turn the call over Aryeh Langer. Thank you, you may begin.

Announcements and Introduction

Aryeh Langer: Thank you Victoria, and as you just heard, my name is Aryeh Langer from the Provider Communications Group here at CMS, and I'm your moderator for today's call. I would like to welcome you to this MLN Connects® National Provider Call on Hospital Compare Overall Star Ratings Methodology. MLN Connects Calls are part of a Medicare Learning Network®.

During today's call, CMS subject matter experts will provide the information on Hospital Star Ratings. This presentation will help you understand the proposed methodology for determining your Hospital Compare Overall Star Rating.

Before we get started with the formal part of the presentation, I have a couple of announcements. You should have received a link to today's slide presentation in an email earlier today. If you have not already done so, you may view or download the presentation from the following URL, www.cms.gov/npc, like in National Provider Call. Again, www.cms.gov/npc. At the left side of the web page, select National Provider Calls and Events, then select the date of today's call from the list. On that page you'll be able to scroll down and open up today's presentation.

Second, this call is being recorded and transcribed. An audio recording and written transcript will be posted to the MLN Connects Call website. Registrants will receive an email when these materials become available.

Lastly, registrants were given an opportunity to submit questions prior to today's call. We thank everyone in advance who submitted those questions.

At this time I would like to turn the call over to Dr. Kate Goodrich; she is the Director of Quality Measurements and Value-Based Incentives Group here at CMS.

Presentation

Dr. Kate Goodrich: Hi everybody. This is Kate Goodrich. I'm very glad to have you on the call today. So I'm going to start with the slide that says CMS Overview, which I think is where we are now.

So first of all, I really want to thank everybody for joining today's stakeholder call for the Overall Hospital Quality Star Ratings Project. We have a very diverse set of organizations and individuals represented on this call that have expressed interest in the star ratings

work at CMS, which I think shows a great enthusiasm for star ratings efforts to improve patient understanding of quality measurement.

The development of the overall star ratings is meant to summarize the existing measures on Hospital Compare and to improve the accessibility of hospital quality information for patients and consumers. As many of you know, we have had many star ratings efforts that have gathered publicity, including Medicare Plan Finder, Dialysis Compare, Home Health Compare, Nursing Home Compare, Physician Compare, and recently, the HCAHPS Star Ratings work.

The star ratings efforts that we've undertaken are a work in progress, and they utilize new methodologies that are both novel to CMS and to stakeholders. So we've conducted this hospital dry run and public comment period to ensure transparency, to solicit feedback from a broad array of stakeholders, and to reflect our flexibility for improvement of this methodology.

Let's go to the next slide for the agenda.

So during this presentation, I'm going to give a short introduction and background to the work so far, the methodology developers will then review the proposed steps for calculating the Overall Hospital Star Ratings on Hospital Compare, and then we'll have some time for Q&A.

Background and Project Overview

Moving on to the next slide for the background. As there's over a hundred measures available on Hospital Compare, some of the information, we know, it can be technical and even intimidating to a lay audience. So we sought to meet the patient call for accessibility by developing an approach to summarize this information as a five star rating.

Star ratings are commonly used metrics both inside and outside of healthcare, and they are easily recognizable by consumers. And patients and consumers have reacted favorably to other CMS star ratings, including our iterative improvements to the methodologies, and we look to continue to improve patient understanding of these measures. So this call is intended to introduce an initial methodology under development. This methodology is not final, and we are actively seeking input on specific aspects from all stakeholders.

Next slide. So as stated before, we do aim to provide patients and consumers with information on many of the multiple dimensions of care that are located on Hospital Compare into a single score. The overall rating is intended to provide patients with a summary reflection of quality and is not intended to replace individual measures that patients will be encouraged to use to understand hospital quality specific to their personal preferences or healthcare needs.

So with this in mind, I'm now going to turn the presentation over to the development team at Yale. And I believe I'm turning this over to Dr. Arjun Venkatesh. So Arjun?

Dr. Arjun Venkatesh: Thank you Kate. This is Arjun Venkatesh, I'm the project lead at the Yale Center for Outcomes Research and Evaluation, and we've been doing work on this development of an initial methodology. And so I want to thank CMS for the opportunity to present this — some of this initial work to the many stakeholders on this call.

This national call really represents one of our many opportunities to get feedback from a broad set of stakeholders. It includes hospitals that are currently part of our dry run, it includes patient and patient advocacy organizations that we have reached out to, as well as purchasers and providers.

Today, we're going to walk through the methodology with some, but not extreme, technical detail. We've got over 2,000 people registered for the call. And recognizing that there's a variety of audiences and a diverse set out there, I want to make sure that we are able to convey some of the key and salient points about this methodology, but keep it at a level for everybody.

We've also received over 30 questions prior to this call. And so I'm going to try to answer a few of those that were common as we go through the presentation. I recognize that many folks will have downloaded these slides, and so I'll try to highlight the slide number as we go through the methodology as well.

We'll start with slide 8, which are the guiding principles for development.

Guiding Development Principles

We embarked on this work here at Yale about a year ago, and at the time, really wanted to set some important principles to guide our development efforts. The five principles that we had set out from the beginning and have used to help guide our work throughout this process are listed here.

First, we really wanted to ensure that our work was towards an end product of simplicity and accessibility. We had originally set a goal of summarizing overall hospital quality into a single star rating. Dr. Goodrich just laid out a variety of reasons why there'd been a call from patients for such work and a need for this accessibility of information. And we have always used that patient's voice in our work, from the initial inception of the idea to the development of methodologies, and now to the development of dissemination materials.

We also sought to convey evidence-based information in a straightforward manner, recognizing that there's many complexities to the individual measures, the types of measures, and a lot of the information on Hospital Compare. We sought to retain those

important scientific properties of those individual measures, while still making them more accessible through a star rating.

Our second principle was to ensure inclusivity. We recognize that Hospital Compare reflects quality across many different hospitals, and our hope with overall star rating was to reflect quality at as many of those hospitals as possible by including as many measures as possible.

Third, we wanted to ensure that this initial methodology was scientifically rigorous. We've utilized established methods in the literature for summarizing scores and we've selected methods very intentionally to maximize the information that we have in our existing Hospital Compare data.

Fourth, we've sought to engage stakeholders. And this has been through multiple channels from the inception of the project to the end. And we'll discuss that in the future. Today is just one example of the many ways we've done that.

And finally, we sought to maintain consistency. We wanted to align our work as much as possible with the existing Hospital Compare website, as well as with a variety of other star ratings efforts that exist throughout CMS. In addition, we wanted to create a methodology that would allow for consistency over time. Recognizing that Hospital Compare evolves the quality measurement evolves, our hope was to develop a broad approach and methodology that could grow as the science of measurement improves.

On slide 9, it outlines a variety of efforts we've undertaken to maximize transparency and engage stakeholders. We recognize that these are new methods and new areas, that as much discussion with the public, with providers, and to get as much feedback as possible not only informs the initial methodology, but any subsequent improvements that we can make.

We started this with establishing a partnership with the patient and patient advocate working group through the National Partnership for Women & Families. These patients and patient advocates have helped inform the development of our methodology, reviewing several of our assumptions across the way, and in helping us think about ways in which this type information could be displayed.

We convened a technical expert panel that was composed of 15 members from a variety of stakeholders. It included providers from the hospital, as well as clinical disciplines. It included patients and patient advocates, it included the purchaser community, and a broad array of other individuals, as well as methodology experts and expert statisticians. That technical expert panel first met in December of 2015. And at that first meeting we discussed the objectives of this work as well as some of the initial principles and processes for selecting measures for the star ratings.

We had another technical expert panel meeting in February of 2015 where we discussed how to account for the many dimensions of quality in a star rating, as well to assess the face validity of some of our initial statistical models.

In the final third technical expert panel meeting that was held in March of this year, we reviewed a revised methodology that incorporated feedback from a variety of those stakeholders, approaches to our methodology being translated into star ratings, as well as thinking about what type of standards would be used for the reporting of star ratings.

In parallel with this, we've also engaged the public through a public comment process. In the winter of this year we had a first public comment period that reviewed the initial objectives of this work as well as the selection of measures for this work. And a second public comment period is currently ongoing that seeks to present this methodology that we present on this call to the entire public for — in seeking of comment.

We're also conducting a hospital dry run so that hospitals can see the methodology firsthand, see what the measures would look like within the context of a preview report, and provide us with feedback as well.

Key Considerations

Slide 10, I think, demonstrates a couple of important considerations that are worth noting before we get into the details of the methodology. We acknowledge that measurement science is imperfect and that it will improve as methods improve as well as the individual measures that underlie the star ratings. And so this initial work is really intended to reflect the quality assessed by the current measures on Hospital Compare. That means that star ratings will evolve as measures are added or removed to Hospital Compare. And that we also acknowledge that existing measures may not capture all of hospital quality.

That said, with nearly a hundred measures of a variety of types that cover many clinical conditions and clinical scenarios, there's a tremendous amount of information available to patients and consumers on Hospital Compare that is worthwhile of summarizing for accessibility. It's also worth noting that the current public reporting requirements result in a large heterogeneity in the number and types of measures reported by different hospitals.

Said another way, there's really no two hospitals that are the same in the U.S. And that's really a reflection of how healthcare delivery is organized. And so, we've been very mindful of that as we've developed this methodology to ensure that we use as much information as possible, that we account for everything we know about every hospital based on the existing measures, but that we also developed a methodology that is fair to hospitals, and recognizing the differences between hospitals, the patients they care for, and the measures they report.

I'm going to stop there for a second because I believe we're going to do a polling, and then we'll go through the star ratings development steps.

Keypad Polling

Aryeh Langer: Thank you Dr. Venkatesh. We're going to just pause for a brief moment to complete, keypad polling, excuse me.

Victoria, can we start that process please?

Operator: Certainly. CMS appreciates that you minimize the Government's teleconference expense by listening to these calls together using one phone line. At this time, please use your telephone keypad and enter the number of participants that are currently listening in.

If you are the only person in the room, enter 1. If there are between two and eight of you listening in, enter the corresponding number. If there are nine or more of you in the room, enter 9. Again, if you are the only person in the room, enter 1. If there between two and eight of you listening in, enter the corresponding number. If there are nine or more of you in the room, enter 9.

Thank you, I would now like to turn the phone back over to Aryeh Langer.

Presentation Continued

Aryeh Langer: And I will turn the call back over to Dr. Venkatesh for the next part of our presentation.

Star Ratings Methodology

Dr. Arjun Venkatesh: Thank you. So we're on slide 11 right now, which is entitled, The Star Ratings Development Steps.

I know the slide may look a little complex at first, but it's really our attempt to try to summarize the entire methodology into a single schematic. We'll refer back to this at multiple points through the presentation. And what we've ultimately tried to do is describe the methodology for overall star ratings as five separate steps.

The first step is the selection of measures. The second step is the grouping of those selected measures into clinically meaningful groups. In the third step, we use statistical models to calculate a score for each of those groups. In the fourth step, we generate a hospital summary score as an average of those seven groups. And then in the fifth and final step, we take the hospital summary score and we use a clustering algorithm to group it into five — a five star rating.

Selecting Measures

In step 1, which is first highlighted on slide 12, I want to begin to describe the selection of measures. This was reviewed as part of a full technical expert panel meeting. It was also part of a public comment that occurred in the winter of this year. And our goal is really to figure — to think about a way to go from a hundred odd measures on Hospital Compare to the defined data set that we thought would be best to use for the Overall Hospital Star Ratings.

We set a variety of principles prior to the selection of measures. We wanted to be inclusive, as I mentioned earlier. Recognizing that there's many different types of measures and many different measures reported by hospitals, our goal was to include as many measures as possible.

The second was to make it feasible. So that we've selected criteria for measures that we knew could be used quarter over quarter, but also ensured that the data was available every quarter in a reliable way for inclusion in star ratings.

Third, we sought to be flexible. Again, this gets at the reality that Hospital Compare information is updated on a quarterly basis. And so, it was important not to set individual specific measure-level criteria when selecting measures, but rather broad principles and broad goals that would grow with Hospital Compare. Ultimately, in April of 2015, which is the period that is used for the hospital dry run as well as the methodology that's presented in our public comment currently, there was about a hundred measures available on Hospital Compare in the downloadable data file. Of those, 75 met the inclusion criteria.

On slide 13, you can see a walkthrough those exclusion criteria. There was up to 106 measures available. Again, this is the same data set that was the downloadable data file for April 2015. I point that out because several of the questions we received prior to the call were around which data set this comes from. And I recognize that for many of the hospitals that are on a call, you may be reviewing data from the recent release in July of 2015. But what's used for the hospital dry run and what is used for this initial methodology that we present is the April 2015 downloadable data file.

And so, of those measures, 11 measures met the first exclusion criteria, which are measures that are suspended from public reporting, retired or planned for retirement from public reporting, or being delayed from public reporting. And there was 11 measures that met that category and were excluded.

Second, we excluded measures that were awaiting future public reporting on Hospital Compare. And the reasoning there was that we should really align every measure that was included in star ratings with the individual measures that patients and consumers could see at the same time on the Hospital Compare website.

Third, we excluded measures with less than or equal to a hundred hospitals reporting performance. This exclusion criteria was chosen because these measures really reflect in measures that were not broadly applicable to hospitals, and so it was felt that at least a threshold of a hundred should be set.

Fourth, we excluded structural measures. And these really come in two flavors. First, we excluded structural measures that are often thought of as measures of registry participation or whether or not HIP structures exist. And second, we excluded volume measures. This is a topic that was heavily discussed by our technical expert panel, that we put out to public comment, and there was broad agreement in both discussions that these measures should be excluded from overall star ratings because of a less of a linkage to outcomes and — is the primary reason.

And finally, we exclude what we term nondirectional measures. Another way of thinking about these are measures for which it's unclear whether a higher score or a lower score is clearly better. One example of this type of measure is OP-9, which is an outpatient imaging measure around mammography use, where the intent of the measure is to reduce variation and not to have the lowest or highest score.

Worth noting is that within this category in the initial version of the overall hospital quality star ratings, we've also excluded the Medicare Spending Per Beneficiary Measure, as clear benchmarks for performance were not yet established.

Ultimately, applying these five exclusion criteria leave us with 75 measures in April 2015 that were included in the methodology and the development work.

I recognize that since this time there are new measures that have evolved and became available since July. Those are not part of our methodology, and so I'd like to avoid that confusion. It's also worth noting that data on some measures may have been suppressed or there may have been other data challenges prior to the development of the downloadable data file that are not included in the hospital dry run. So the data set used is really the downloadable data file that was initially available for April 2015 for this hospital dry run.

Standardizing Measures

Slide 14 describes what is really the final analytic portion of step 1. Once measures are selected, they needed to be put into a common language, or I guess you should say a fashion suitable for combination. And so the first thing that we did is that we recognized that the quality measures on Hospital Compare really include a wide variety and type of measures. There are measures of time, such as emergency department throughput measures. There are measures of percentages; these are often process measures, such as anticoagulation for stroke patients. And there's measures that are reported in rates, such as readmission rates or mortality rates.

We wanted to put all of these measures into one common scale. To do that, we standardized these measures along the normal standard distribution or by calculating the Z-score for each measure. In doing that, we put every measure on a curve, not necessarily a normal curve, but a curve that ensured that higher performance was associated with a higher Z-score and lower performance with a lower Z-score.

By that, what I mean is we flipped the direction of measures that may — for which a lower score may be better. So mortality, for example, a lower mortality rate is a better score. But in the context of star ratings so that we could combine all these measures, we flipped that so that we — a higher score is better.

To avoid any confusion while these are used internally within statistical models, we don't report and we try to avoid reporting any numbers that would cause confusion by seeing a positive number where we're used to seeing a lower number usually.

Finally, after putting all these numbers on a common scale and getting them all in the right direction, we noticed that there are a few very extreme outliers. And those extreme outliers, what we did with them is a process called Winsorization. And so what we mean is, if a score was above the 99.75 percentile, we reduced it to the 99.75 percentile. And we did the same on both ends of the distribution.

This affected less than 1 percent of hospitals and less than 1 percent of measures. And it's worth noting that these changes that I've described here to prepare the data for overall star ratings really have no material impact on hospital measurement. We explain them so people understand how the data was prepared and put together before the statistical models were applied.

Measures Grouping and Latent Variable Modeling

Slide 15 shows step 2. Once we had our 75 measures for this April 2015 period together and we had them in the same direction and we had them on common scale, it became important to group these measures.

There was a variety of principles used for this grouping that are listed on slide 16. And it was really from a discussion that happened amongst the technical expert panel as well as from our previous literature review that acknowledges that hospital quality is multidimensional. There's a variety of frameworks that are used to define those dimensions, and it's unclear whether one framework vs. another framework is a better way for grouping these measures. But what we tried to do is balance what was available in the literature and what we knew about frameworks for our grouping measures of quality with the reality of what measures currently exist on Hospital Compare.

And ultimately, what we decided on was using seven measure groupings. The seven measure groupings we used are aligned with the current Hospital Value-Based Purchasing Program, so it met that initial principal of consistency that we had discussed

early on. The seven measure groupings that we used are current categories on Hospital Compare. And so in doing so, it allowed for consistency between what patients and consumer see on the current website and what they would see in star ratings. And third, these measure groupings are also consistent with a variety of other national quality initiatives, as well as language within the CMS quality strategy.

When we presented these groupings to our technical expert panel, they felt that they were clinically reasonable, that we had selected measures and put them within groups that made sense. And finally, these groupings were — met that criteria flexibility I mentioned earlier, that they would allow us to add future measures or remove measures as Hospital Compare evolves.

On slide 17, you see the seven measure groupings. Three are groups of outcome measures. The first is of mortality measures; there are six of those measures included in star ratings. The second outcomes group is of measures of safety, and there are eight measures within that group. The third group is measures of readmission, and there are seven measures included in that group.

The fourth group is of patient experience. At this point in time that includes 11 measures that are all from the HCAHPS program.

There are two groups of process measures. One is of 30 measures of effectiveness of care. These are measures of adherence to evidence-based guidelines. The second is a group of timeliness of care. These are frequently measures of time and quickness of care.

And finally, we have a grouping of efficiency measures, which currently is comprised of imaging efficiency measures, which are part of Hospital Outpatient Quality Reporting Program.

I'll pause for a second because a few of the questions we received prior to this call were around what the difference between the Overall Hospital Star Ratings project was and the HCAHPS Star Ratings, which went live in April of this year. Much of it is attributed to this patient experience group.

The 11 measures included in the Overall Hospital Star Ratings are the identical 11 measures used in the HCAHPS Star Ratings. In fact, we used the same linear mean score, which is the adjusted score for each hospital and each question that's used as part of HCAHPS Star Ratings. In addition, the HCAHPS Star Ratings requires a minimum of a hundred surveys be completed. And the same minimum threshold is used for inclusion in the Overall Hospital Star Ratings.

The different slides in step 3, which is in the way that that data is combined into a score for inclusion in the Overall Hospital Star Rating. I will say that we've conducted

numerous analyses assessing the comparability between the two and find — found a concordance greater than 99 percent between the two methodologies. And so ultimately it should have no material impact on a hospital's patient experience score.

If you look at slide 18, it shows step 3, and so is a reminder of where we're at. In step 1, we selected measures from all of those available on Hospital Compare and we put them in a common score and a common direction. In step 2 we grouped those measures by the seven measure groups. And now in step 3, we apply statistical models to calculate a score for each of those groups.

Slide 19, describes this approach. We selected to use latent variable models for each group to combine the scores. Latent variable models are an analytic approach that seeks to measure dimensions of quality, such as hospital safety, that can't be measured directly. There's no single measure where we say absolutely captures all of hospital safety. But we recognize that we have many individual measures of hospital safety, that each tell us something about that latent or underlying overall hospital safety. And so we estimated based on the exiting measures we have and what we can see.

A separate latent variable model is used for each of these measure groups. So there are seven statistical models total, one for each of the seven groups. These latent variable models have a variety of advantages, and they were selected because they can accommodate missing information, they can accommodate diverse hospital reporting patterns, they accommodate the addition and removal of measures over time, and they consider the relationship between measures. I explain this a little — I'll explain this a little bit more on slide 20, which goes through some of the advantages and challenges with these models.

The latent variable modeling approach is a method that's been used for composite measurement in the healthcare quality literature. There's multiple references of its use, and it really goes back to that scientific rigor principle that we established early in our work.

Second, this modeling approach allows us to account for the consistency of performance by giving more importance to measures that are more highly correlated within a measure group. While we may know that within a context to readmissions we have six separate measures, it may be that one of those measures and a few of those measures tell us more about that latent quality around readmissions that we can't capture with a single measure, and this approach lets us do that.

Third, the approach allows us to account for missing information. As I mentioned earlier, there's highly variable information based on the reporting profile of each individual hospital. Latent variable models only give credit to hospital performance based on information that exists without punishing hospitals when information is not present. And that's really an advantage when have such diverse hospitals across the country. And

finally, by using a modeling structure, we're able to account for the sampling variants or the number of cases we have for each hospital included.

One of the challenges with this approach is it could be challenging for patients and consumers to understand. We've presented this to our working group of patient and patient advocates and found that, while at first it — there is some initial resistance, that it actually was well received and that the methodology overall seemed to make sense when explained.

Second is that each latent variable model assumes that each group reflects a distinct aspect of quality. And so if there are measures that we've assigned to one group that also tell us about quality in another group, we can't account for that in this current methodology.

And third and finally is that we've assumed in our work that each quality measure included is a valid indicator of quality. We sought to do this work based on existing data that was already in Hospital Compare and to not reevaluate each individual measure. And in doing that we've included — been as an inclusive as possible.

Slide 21 describes how we've accounted for sample variation. For each measure, hospitals report a different number of cases, and we can account for this variation. Hospitals with a larger denominator or more cases included may have a more precise score. And we've accounted for that in the model by using a weighted likelihood.

Slide 22 introduces a concept of measure loadings. This is probably a new term to most who are on the call. And what a loading is, is it's the degree to which an individual measure is associated with that hospital-specific measure group's score. Said another way, it's the extent of that measure's association to a latent aspect of quality relative to or in comparison to all the other measures within that group. This is calculated — or is estimated by the model and not set by us individually.

These measure loadings are the same for every hospital. And what that means is — so a given measure's loading, it's the same loading for every hospital in the data set. And measures that have higher loadings are simply measures that are more strongly associated with the group's score or the latent aspect of quality. An example of this may be within the readmission measure group, where the hospital wide readmission measure has the highest loading and is therefore —seems to be the most associated with the readmission group score.

Large measure loadings, however, don't imply a weight, and it doesn't necessarily mean that that one measure matters more. It's very possible that a measure with a higher loading has high performance across many hospitals. And if there's little underlying variation, it wouldn't mean that that single measure drives the group's score. Ultimately

this allows us to include multiple measures at a time while accounting for the consistency of performance across multiple measures.

Calculating the Hospital Summary Score

Step 4 is the next step. At the end of step 3 we have seven statistical models that have generated separate group scores for each hospital. Step 4 seeks to combine those group scores into a hospital summary score. It's described on slide 24 how we do this.

It's a simple weighted average. The hospital summary scores are first calculated for each of the seven groups. These seven groups have a weight that has been adapted from a Hospital Value-Based Purchasing Program and after — with input from our technical expert panel, and it's currently in public comment. The three outcome groups are each given a 22 percent weight, as is the patient experience group. And the two process groups and the imaging efficiency group are given a 4 percent weight. Based on the available measure group scores, a weighted average gives us the hospital summary score.

Slide 25 describes the principles we used for developing these weights. We considered the importance of these measures, not necessarily from an individual measure perspective but as a group of measures and in their consistency with national priorities and policy objectives. We've emphasized outcomes as such as well as patient experience. We've also solicited a lot of stakeholder input, both from our technical expert panel, our working group of patient and patient advocates, and now currently through both the hospital dry run and the public comment period.

We conducted a survey of our TEP members that also informed this weighting of groups. And ultimately, we found that there was strong consistency between each of these sources of information, where most felt that outcome groups should be emphasized and the relative equality of those were also valued.

Slide 26 describes how these weights compared to the fiscal year 17 Hospital Value-Based Purchasing weights. These are policy-based weights, and you can see they're fairly consistent, but there are differences between the two that required differences in weights. We're currently seeking comment on this. This is an aspect of the methodology that is certainly open to revision. And so, for those that are hospitals participating in the dry run or other organizations that are participating through public comment, we encourage you to send feedback to our inbox regarding these weights.

Slide 27 describes what we would do when not all groups are present, and this happens because we recognize that not all hospitals report all 75 included measures. If a hospital has no measures in a group, we consider that measure group missing. So they may have six groups or five groups of the total seven.

Our development team has recommended using the same approach as the Hospital Value-Based Purchasing Program in this instance by redistributing weight to the nonmissing measure groups. This maintains the proportionality of weights so that outcome measures are just as important in comparison to process measures and available outcome measures are still proportionally equal.

Assigning Star Ratings

Step 5 is the final step. And this is to assign star ratings and is shown on slide 28. Our goal here was to use a clustering algorithm to go from an individual hospital summary score for every hospital into one of five star buckets.

Slide 29 describes this. The hospital summary scores are classified using an algorithm called k-Means Clustering. Statistically this is defined as determined by the sum of the squares of distance between the hospital summary score. Another way of thinking of this is that five separate means are set along the distribution of summary scores, and a hospital is assigned to one of those five clusters in a way to maximize its similarity to the other hospitals in its cluster but to ensure that it's most dissimilar from the adjacent hospital cluster.

So a hospital that may be between four stars and five stars, if it's assigned a five star, it's most like the five star hospitals and less like the four star hospitals. You can see the description of the five star categories there, where three stars were set as the average performance on the hospital summary score, four was set as higher than average, and five is the highest and is symmetrical on the other side.

Slide 30 describes some of the advantages and challenges of the k-Means Clustering approach. The advantage is that it enables the designation of five means, which matches well to the five-star objective of star ratings. It also allows and ensures that hospitals in a cluster have similar and meaningful summary scores. Third, the approach also produces a slightly broader distribution of star ratings than some alternative approaches we evaluated. And finally, in testing that we have done, we found largely pairwise statistically significant relationships within each of the measure groups. Meaning that in the case of mortality, five star hospitals do better than four star hospitals, that do better than three star hospitals, that do better than two, and so down.

There are few exceptions to this, which should be expected by combining several dimensions of quality, but it largely reassured us that the methodology developed was separating statistical significances.

One of the challenges of a k-Means Clustering approach is that the majority of hospitals fall into the three-star cluster. This is not unique to the overall star ratings, and it's common to many aggregate or composite measurement strategies, given that there are many hospitals that often perform near the middle or around average performance.

Second, this approach could be difficult for patients and consumers to understand, and we're seeking as much public comment as well as comment through the dry run to think about ways in which we can communicate this most effectively.

Slide 31 shows the distribution of star ratings for hospitals that met reporting thresholds in April of 2015. As you can see, 70 percent of hospitals were at the three star, about a little over 500 hospitals were at two and four, and then a small number of hospitals were at one and five stars nationally.

These numbers are meant to just reflect how the methodology played out in April 2015. And we recognize that in a different quarter of data the star distribution can evolve as both the measures change and hospital performance on those underlying measures change.

Star Ratings Thresholds

Slide 32 describes the star ratings thresholds. As we develop the methodology and presented it to our technical expert panel, it became clear that some hospitals report very few individual measures. And while the summary scores are calculated using all available data and all available measures, we know that as the hospitals have fewer and fewer measures, the reliability or the face validity of assigning a star rating may be less so. And to the development team recommended setting a minimum reporting threshold. This is similar to what we do from many individual measures as well as what —to what is done for the Hospital Value-Based Purchasing Program. This threshold was based on both reliability calculations as well as the assessment of face validity with our TEP.

The current reporting threshold that is included in the hospital dry run was to require at least three of the seven measure groups, of which one had to be an outcome measure group, meaning mortality, safety, or readmission, and at least three measures in all of the included groups in order to get a star rating reported. Again, we're seeking input on this. And as I describe on slide 37, that can be submitted via the public comment or the hospital dry run — slide 33, sorry.

So on slide 33 it shows how this decision was made. As you can see, along the rows there are different numbers of minimum measure set per group, and across the columns the number of minimum groups that a hospital would have. The threshold that was set for this hospital dry run was three and three — that's the highlighted box. And what that indicates is that 3,700 or 78 percent of all hospitals on Hospital Compare met that threshold for reporting.

If we were to loosen that threshold and only require two groups, that would increase to 84 percent, or if we were to make it more strict and require four groups at the same three measure threshold, it would reduce the number of hospitals that had a star rating

to 70 percent. This is largely a policy-based decision that, as I mentioned, we seek input on from the public.

Dry Run Timeline and Resources

Slide 34 describes our current our timeline. The current star ratings dry run started on July 17th. Many hospitals have already downloaded their preview reports and seen their individual results. It will continue until August 17th. During this time, we're asking that stakeholders send any questions or comments to the inbox that you have in your slides.

There's also additional resources for the dry run available on QualityNet that includes not only a Hospital User Guide to guide hospitals through their individual report on star ratings, but it also includes a methodology report that details this in more technical detail, including technical appendixes. We do make one request, that no personal health information, or PHI, is included in the comments or questions that you submit to the inbox.

At the same time, as you can see on slide 35, we are also holding a public comment, that's also going from July 17th to August 17th. And we're encouraging stakeholders to submit comments to this public comment through the CMS Quality Measures Public Comment Page. Again, we ask that no PHI be included in those comments. And we really do appreciate any feedback that we get both during today's question-and-answer period, as well as through both of these vehicles as they can really make for some meaningful improvements to our methodology.

We'll now move over to the question-and-answer session.

Question-and-Answer Session

Aryeh Langer: Thank you very much Dr. Venkatesh. Our subject matter experts will now take your questions. But before we begin, I'd like to remind everyone that this call is being recorded and transcribed. Please state your name and the name of your organization once your line is open. In an effort to get to as many participants as possible, we ask that you limit your question to just one. Victoria, we are ready to take our first question, please.

Operator: Certainly. To ask a question, press star followed by the number 1 on your touchtone phone. To remove yourself from the queue, please press the pound key. Remember to pick up your handset before asking your question to assure clarity. Please note your line will remain open during the time you are asking your question, so anything you say or any background noise will be heard into the conference.

Please hold while we compile the Q&A rooster.

Your first question comes from the line of Millie Russin.

Millie Russin: Yes, Millie Russin here from UF Health Shands. My question pertains to the one thing that you said people may not be terribly familiar with. You mentioned that measures with higher loadings are more strongly associated with the group score, but that's not reflecting the higher loading for the hospital specific — it's not identifying key drivers for the hospital. Am I correct in understanding that?

Dr. Arjun Venkatesh: Yes. So I think that — it's a good question. I know that this is — it's probably one of the most confusing parts of the entire presentation, the methodology as a whole. And I'm going to kind of break it into a couple parts.

I think first, it's worth noting that the use of the statistical model is really to allow us a way to reflect quality by combining the individual measures and to put some structure on some data. And so what that means is that it was really developed and selected as an approach so that we could combine information and make it accessible to patients and consumers. And it wasn't necessarily designed to say that we wanted to identify how to drive quality improvement at the individual hospital level.

I think that from the perspective of guiding hospital QI, it's really the individual measures that already part of Hospital Compare that are more valuable for that. And this is really just kind of a glance or summary of those.

That said, for each individual measure loading, it's not necessarily an individual weight to say that that measure, if you improve in that one measure that had the highest loading, say, within a group, would therefore result in a higher measure score, and, therefore, an overall higher star rating. And the reason for that is that the performance on measures will change over time. And so just because a measure is loaded more in April of 2015, it's possible that as new measures are introduced, and at the performance of hospitals change on the underlying measures, that the same measure may not be loaded the most in July of 2015.

Now, it's most likely that measures that are loaded more in each period will retain and be the higher loaded measures, maybe it's not the highest. And the other thing worth noting is that just because a measure is the highest loaded, there may be little bit — there may be lower variation for that individual measure. It doesn't necessarily mean that that's a high variation measure with a lot of impact on the summary score just because of that.

And so, the guidance we have said, and what we've said to hospitals and the Hospital User Guide that comes with this, is that the star rating is really a reflection of the existing measures for patients and consumers and not really intended to guide specific quality improvement efforts. And that the loadings are a way in which we can give more credit to measures where there's consistency of performance. And so that measures that move together within a group are given more credit because we believe those reflect that latent aspect of quality we can't measure as well.

Millie Russin: That's very helpful. Thank you.

Operator: Your next question comes from the line of Deborah McDonough

Male: Hello.

Aryeh Langer: Hello Deborah.

Male: Thank you for an excellent discussion and, having worked in the CMS QAO program for two decades, please say hello to Harlan, and thank you for the work.

If you publish the scoring ranges for the five bands, in the download report, the scores are there but the cut points are there, or is the distribution like it is because you've set up a forced standard distribution with cut points and how you decide the cut points?

Dr. Arjun Venkatesh: Thank you. We'll be sure to say hello to Harlan, and thanks for the question.

So we didn't actually set any cut points. The cut points that you see in the downloadable methodology report are the cut points that were generated by the k-Means Clustering algorithm. And so what the k-Mean Clustering algorithm starts with is whatever the underlying distribution is of the hospital summary score. It does not assume a normal distribution, and the distribution itself is well, roughly normal, it does have a slight tail to it. And what it does is, that the k-Means algorithm is essentially an optimization technique, where it tries to set five means along that distribution in order to allow for the clustering of similar hospitals. It then outputs those cut points, and so those were the cut points that we reported for April 2015.

If we were to repeat this in July of 2015, let's even pretend the measures didn't change but underlying hospital performance did evolve and change on those measure, then you would have different group scores, which would result in a different summary scores, which will lead to potentially a different summary score distribution, where if we ran the k-Means algorithm again, we would get different cut points.

Male: Right.

Dr. Arjun Venkatesh: And so it wasn't that we set any line in the sand to say that is five, four, three, two, one, but rather we allow the clustering algorithm to sort between and set those bands.

Male: Great. All right, well, thank you. If I could just provide one comment about that, if you could make that clearer in the documentation because, although I'm familiar with

the technique, I didn't pull that out of the information, and I was left wondering the question I had. But thank you very much.

Dr. Arjun Venkatesh: Thanks, will do.

Operator: Your next question comes from the line of Nancy Foster.

Nancy Foster: Hi, good afternoon, Nancy Foster with the American Hospital Association. Thanks for providing this very helpful and educational session. I wanted to ask a question about the minimum numbers — the minimum measures and the minimum number of measure groups that are reflected on page 33 of the document.

I'm a little concerned that this small number of measures and small number of measure groups may create some real oddities in the data. And, you know, in fact, as we look at the measures that you currently have to work with, they don't reflect a lot of the care in hospitals at all, nothing on cancer care, nothing on maternal and child health to speak of really.

So as we're thinking about trying to create something that honestly reflects the quality of a hospital overall to the consumers who want this information, have you looked at the data to see if there are aberrations created by selecting this small number of measures and small number of measure groups?

Dr. Arjun Venkatesh: Thank you Nancy, a lot there to include in that I want to try to cover.

I think, when it comes to minimum measures, probably, the best place for folks to look and to think about that topic is page 25 of the methodology report that was part of the hospital dry run. And when we came up with those minimum thresholds, we did it via a couple angles. One angle was more technical, and so we tried to calculate statistical reliabilities and tried to determine how many measures we needed within each of those groups to hit a different reliability threshold. And we set the threshold very high.

We initially said — initially said we'd like to be above 0.7, and what we found then was that two measures actually within each group would give us statistical reliability for each of the latent variable models for each group. When we set the measure threshold at three, which is where it is during the hospital dry run, it exceeds 0.75 for all of the measure groups.

And so from a technical perspective, we feel like we have a lot of reliability within each measure group. But that's not all of it, and I think that, you know, that has to be interpreted, like you mentioned in your comments in the context of the measures, the groups themselves. And so we showed this to the technical expert panel to get some of their feedback as well. And people felt that, you know, looking at the reliability

calculations, looking at the measures, and looking at that table I showed earlier about how changing the thresholds impacts the number of hospitals with a reported star rating, that that minimum measure count of three seemed reasonable.

And like I said, it's a place we're certainly open to comment on it. If folks feel like it should be higher and that the feasibility isn't there at three, that's the comments you want to hear in the inbox.

The other aspects of what you asked about, I think, is more related to, you know, the underlying measures. And the scope of work we had was really confined to the existing measures on Hospital Compare. And so that was kind of the starting point we had. And we recognize that it doesn't include, as we've said before and as Dr. Goodrich also discussed, it doesn't encompass all hospital quality — is there's quality measurements and enterprise that's still growing and will evolve over time as we get more measures that cover more conditions and more patients. And so as we developed this methodology based on the existing measures, but with that kind of eye towards growth, or the flexibility for the future.

I think anything else about the specific measure I'll leave to CMS, as it was kind of outside the scope of our work.

Dr. Kate Goodrich: So Nancy…

Nancy Foster: Thank you, and I do get it, it's sort of like trying to create a jigsaw puzzle from a bunch of disparate jigsaws that has been thrown on the table so ...

Dr. Kate Goodrich: Yes.

Nancy Foster: It's interesting, it's challenging. I appreciate your effort.

Dr. Kate Goodrich: Nancy, this is Kate, just to chime in a little bit. I think Arjun's answer was perfect. You know we did — I do want to emphasize that we did design this as Arjun said, with an eye towards the future in the hopes that over time, the measure portfolio can grow to more fully reflect, as you say, sort of the – as much of the totality of quality of care as we can. I'm not sure we'll ever get to a 100 percent of the totality of quality care, but we certainly do recognize that need. And we think that this — we hope anyway that this methodology can be flexible enough to allow that to happen.

Nancy Foster: Thanks Kate. And thanks for letting us comment along the way.

Operator: Your next question comes from the line of Priscilla Magen.

Rory Jaffe: Yes, this is Rory Jaffe from CHPSO Patient Safety Organization. I have a question about using as input some measures that were originally designed for

reimbursement or quality ranking, for example, the issue with readmissions rates and not using socioeconomic status adjustments for that appears to significantly penalize those hospitals that serve the underserved and rank them lower as a result.

Dr. Arjun Venkatesh: Thanks for that question. And so, I think this is in some ways related to the previous answer, which is that when we embarked on this work we really set some of those initial principles of inclusivity upfront. And in doing so we wanted to be as inclusive of all the measures we had on existing — that were existing within Hospital Compare. And so and, that was also kind of agnostic to each measure's use within a payment program. And so some of these are used in payments, some are not. Some were probably not used in payment at one time and may have evolved to be used in a payment program. And so, in many ways, the overall star ratings methodology has been developed really with that patient's voice in mind upfront, which was to include all available information and agnostic of the payment side of it.

With respect to the socioeconomic status, I think, it sort of related the fact that we know we developed this methodology with the existing measures. And that's a work that is underway as a result of the IMPACT Act. It's ongoing work at NQF, and so these measures will evolve and change over time. And as they do, that would be reflected, those would be the same input that would go into the methodology. But at this point, we felt that because those measures are on Hospital Compare and because those are the individual measures the patients would see, that it was important to include anything a patient would see as an individual measure as much as possible in the star rating.

I'll leave it to CMS in case they have anything they want to add as well.

Dr. Kate Goodrich: This is Kate Goodrich. Just to add, the SES issue that you raise is certainly a very important one, and there is a lot of work going on in that sphere. And so should the measures change over time, that would absolutely be reflected.

Rory Jaffe: Yes, and it's kind of based on my concern that these measures are seen equally once they pass that initial disqualification phase. And some of these measures are much more problematic than others. I mean once — everything you do past there is very interesting and very sophisticated. But it's the inputs themselves that seem to be problematic and will be giving star ratings that may or not – may not properly steer consumers towards better hospitals.

Dr. Arjun Venkatesh: Thanks, we appreciate that. And I think if there's, you know, any specific measures that you really have comments on that you feel maybe we should rethink or criteria, that would be the place where we'd really appreciate that feedback in the inbox as we review the methodology and think about revisions.

Rory Jaffe: Thank you.

Operator: Your next question comes from the line of Carlene Parrish.

Carlene Parrish: Hi, this is Carlene Parrish with Hoag Orthopedic Institute in Irvine. Once we have a published summary star rating, will a separate star rating value continue to be published for a patient experience on Hospital Compare or will that go away?

Kristie Baus: Hi, this Kristie Baus from CMS, I am the Hospital Compare team lead. And at this point in time, once we've determined whether or not we're going to move forward with the overall star ratings, the HCAHPS Star Rating will in fact continue to display as a separate, distinct star rating.

Carlene Parrish: Great, thank you.

Kristie Baus: You're welcome.

Operator: Your next question comes from the line of Lana Webel.

Lana Webel: This is Lana Webel at DeTar Hospital in Texas. My question is, when do you anticipate if everything goes, that it will go live on the Hospital Compare?

Kristie Baus: And hi, this is Kristie Baus again. We are still working through those details. If we do determine to move forward with the star rating, the hospitals will receive a preview report prior to any public posting of their data.

Lana Webel: Thank you.

Operator: Your next question comes from the line of Doug Johnson.

Doug Johnson: Hi, this …

Aryeh Langer: Go ahead.

Doug Johnson: Hi, this is Doug Johnson, University of Missouri, and the lady just ahead of me, beat me to the question. Thank you.

Aryeh Langer: Thank you.

Operator: Your next question comes from the line of Mohamed Sylla.

Aryeh Langer: Your line is open.

Operator: Mohamed, your line is open. Please proceed with your question.

Aryeh Langer: I think we should move on to our next question, please.

Operator: Your next question comes from the line of Amy Chin.

Amy Chin: Hi, this is Amy Chin with Greater New York Hospital Association. So my question is we – some of our hospital — member hospitals received their dry run report. And when trying to validate the results, did you use the actual extracts from Hospital Compare to drive your models or did you use a different version of the data?

Dr. Arjun Venkatesh: So we used the downloadable data file for April 2015. That said, the current downloadable data file may not be sufficient to do a full validation. And the reason being is that we also wanted to account for sampling variation. And so the downloadable data file would have all the individual measure scores for every hospital that we included, and that's the same score that we used. But it would —may not have is the denominator or the sample size for every hospital. It'll have it for some measures but not all. And so we were in the process — and we went through a process of obtaining that for every included measure in order to be able to actually construct the statistical models. The intention would be that if this were to move forward, that all that data would be available together in one downloadable data file.

Amy Chin: And so part of that question would be would — how would hospitals validate their data if not all the data is available to them?

Dr. Arjun Venkatesh: Sure, I'm going to lead — I'm going to allow CMS to comment. I guess the only thing I would say first is that, you know, it'd be — I guess in some ways similar to some individual measures that exist out there where they're centrally calculated where the individual hospital data is not necessarily all available but you need the data from every hospital in order to actually run the statistical model. Our intention here though is to actually use data that are largely publicly available and to be as transparent as possible. And so, I'm going to leave it to CMS, but I think in this case, for something like this, this type of information could be made available.

Kristie Baus: So hi, this is Kristie Baus again. And just, you know, thank you for your comment. We do appreciate any feedback. So if we move forward with the star ratings and you will be receiving a preview report for your hospital. And that preview report should contain enough information that will help you understand and to see where your star rating came from.

Again, we're kind of like in the process of finalizing the methodology right now. So, I mean, your comment really helps because it lets us know that hospitals are looking for more information. So, thank you.

Amy Chin: OK, thank you.

Operator: Your next question comes from the line that Mohamed — I'm sorry he has withdrew his question.

Your next question comes from the line of Tara Engstrom.

Tara Engstrom: Hi, I wanted to find out if you were going to include behavior health care facilities in the star ratings?

Dr. Kate Goodrich: For this particular project, we are only doing the acute care hospitals. We have not incorporated any behavior health or inpatient psychiatric facilities at this time.

Tara Engstrom: OK, thank you very much.

Dr. Kate Goodrich: You're welcome.

Operator: Your next question comes from the line of Gloria Beecher.

Gloria Beecher: Hi, this is Gloria from Grady Health System. Thank you. My question was asked by Nancy. I had a question regarding the measures. But it was duly answered. Thank you.

Operator: Your next question comes from the line of Rex Graff.

Rex Graff: Hi, this is Rex Graff from Munson Medical Center in Traverse City, Michigan. I'm excited about this. And if there are changes to the methodology, will there be another dry run preceding a preview report? And also, when they are published publicly, will there be accompanying rules or guidance about how to use this for hospital transparency like marketing or can we put it up in the hallways kind of thing? Will there be rules or guidance published to accompany them?

Kristie Baus: Hello, this is Kristie. That's a very good question. So as far as the marketing goes, the only guideline that we have for the use of the data, you know, that it is publicly available data and we just ask that if notated that data does come from CMS. So you can put post, you know, posters in the hallway if you'd like.

And as to the first question, can you repeat your — the first part of the question, I apologize?

Rex Graff: There's already been a delay in working on these. If there are changes to the methodology, will we go through another dry run where we can go through it with you or will it just be a preview report of we changed it, here's what's going to be published.

Kristie Baus: So, thank you. The purpose of the dry run is to help CMS kind of work out any glitches we might have in the event that we do decide to go live. At this point in time there are no future plans for another dry run. You know, you will be receiving a preview report if this project continues on. And that preview report will also contain

specific guidance to help hospitals better be able to understand the preview report and the data that's included in it.

Rex Graff: Thank you very much.

Kristie Baus: You're welcome.

Operator: Your next question comes from the line of Tina Charbonneau.

Tina Charbonneau: Hi, this is Tina Charbonneau from the Adirondack Medical Center. And my question has to do with some of the measures that are on Hospital Compare that are indicated only as no difference than national average. And so how do you do your calculations when the values are no different than the national average, and thanks?

Dr. Arjun Venkatesh: Thanks Tina, that's a good question. The way we input the data into the models is we actually use the numerical scores. And so, while many measures may be displayed on Hospital Compare as better than average, below average, or no different than the average. That's really the kind of categorical summary of that numerical score to make it accessible to patients and consumers for an individual measure. But underlying that is a actual percentage or an actual readmission rate or an actual number. And the input, the measures that we use up front, in what I describe as kind of step 1 of the methodology, is based on the actual number. So we take the actual numerical score, we standardize that so it still is a number, and we put all in the right direction before we include them.

Tina Charbonneau: OK, thank you very much.

Operator: Your next question comes from the line of Laura Sodemann.

Greg: Hi, this is Greg from Waukesha Memorial Hospital. Along the same lines as Amy's question, we're trying to validate the data that we have available from the five star group for the data sets off of Hospital Compare for April, and we're having some difficulties matching the data we have from you guys and data that's on Hospital Compare. What recourse do we have when we see inconsistencies?

Dr. Arjun Venkatesh: Sure. So I think for specific questions like that, I think the easiest way to make sure that we get your question answered is if you can email the Dry Run Inbox that's listed. And make sure to include you're hospital number, that helps us help you resolve any of the discrepancies you may see between the data and what was reported in your preview report.

I will say that one of the common reasons there's a little misunderstanding is that you already mention that you looked at April. Some hospitals have looked at July and not

realized it was April. Another has been that for some hospitals the standardization of the score results in a number that looks different than what you're used to, and so that may be the other reason. But if you email us in inbox, we're happy to get back to you and help resolve any potential discrepancies.

Greg: Yes, we are kind of on a timeline, we need to be completed by the 17th, and there is a 3-day turnaround time for the CMS Star Ratings site. Is there a way to expedite this?

Dr. Arjun Venkatesh: We will — any question that's comes into the inbox, we will be sure to answer. It doesn't have to be finished by the 17th. As long as you can submit the question, we will connect with you. And even if you're able to just submit us part of it, we can kind of go back and forth to make sure that we address each of your questions.

Greg: Great, thank you very much.

Operator: Your next question comes from the line of Tiffany Miller.

Tiffany Miller: Hi, this is Tiffany Miller from Neosho Memorial in Kansas. And my question really is about — we're a critical access hospital, so we have low volumes but great results. So for example, on the outcome safety measure data for, just for example CODI and CLABSI because our utilization is so low, our infection rates are also low. But that data's not being captured because of the low volume. So it kind of seems like we're being penalized for good results.

Dr. Arjun Venkatesh: Thanks Tiffany. And so, as I understand the question, just to make sure I'm getting this right, it's about the fact that because you're case counts are low, you don't get to — you don't have a publicly reported score for the measure. Is that where this comes from?

Tiffany Miller: That's correct.

Dr. Arjun Venkatesh: OK, yes. So this is — we understand that. And we recognized this from the beginning of the development process when we looked at all available measures on Hospital Compare right now. We recognize that hospitals have very different reporting profiles. And so what we sought to do was first pick a statistical model and an approach that didn't seek to penalize hospitals that have fewer measures.

Now — so what that means is that it's not set up so that only large hospitals with lots of measures can only do well, but rather that we only give credit to hospitals based on the measures they have. And so if they perform well on their fewer measures, then they will do better. If they perform poorly on their few measures, they would do worse. And so we've looked at that a little bit. And we're going to keep looking at it. But in our first initial glances of looking at things such as based on critical access hospitals or hospital size, we're finding that there's a broad range of performance across these types of

characteristics. And so there are hospitals that fit the full range from one to five across these different types. And so I think that helps us feel more reassured that there's not like a penalty for the smaller hospitals.

Now that said, I think the one thing worth noting is that the use of these statistical models to look at hospital performance, because they don't penalize a hospital for having fewer measures, it sometimes also means that we don't have a lot of information about hospitals that may be smaller. And so there's a possibility that very— not small hospitals, but hospitals that report very few measures, I should say. And by that it could be because they choose to not report or could be, like you said, because they don't meet the minimum case counts. There may be statistical phenomenal of shrinkage where they go more towards the middle and they're more likely to look like they're average because we have fewer measures from which to assess their performance. But like I said, when we tried to see if that plays out, it doesn't appear to be that way initially. It appears that there's a broad range of performance even for smaller — or for hospitals that report fewer measures.

Tiffany Miller: Thank you.

Operator: Your next question comes from the line of Melanie Hoover.

Melanie Hoover: Hi, this is Melanie Hoover from Magruder Hospital in Port Clinton, Ohio. And we also are critical access and with — being critical access, we're not part of the Hospital Readmission Reduction Program from the Affordable Care Act. But with this methodology, our readmission would be factored in, is that correct for the outcome?

Dr. Arjun Venkatesh: Yes, so this methodology, like I mentioned earlier, it's not tied to the CMS payment program, so things like the Hospital Readmission Reduction Program or the HAC program are really unrelated to the star rating.

What the star rating is based on is whether or not a hospital pledges for public reporting in either the Inpatient Quality Reporting Program, IQR, or the Hospital Outpatient Quality Reporting program, OQR. And so if your hospital pledges in those programs, then those measures which are the ones that are the same as those that are publicly available on the Hospital Compare website are the ones that get included in the star ratings.

Melanie Hoover: OK, thank you.

Operator: Your next question comes from the line of Beth Dibbert.

Beth Dibbert: Hi there, Beth Dibbert from Rural Wisconsin Health Cooperative. And I appreciate your response a couple of questions back regarding calculating methodologies and models for critical access hospitals and other low-volume hospitals.

But I'm curious about how you're assured of your methodology and your statistical model when the April file that you used to generate the HSR reports, at least in our experience with Wisconsin Critical Access Hospitals, a large number of the process effectiveness measures were not available in that database.

So I'm curious about that. And I'm also curious to hear your answer again about, if this error replicates itself, what recourse do hospitals have in the next dry run? So thanks for your answers.

Dr. Arjun Venkatesh: So I'll answer part of that, and then I know CMS will want to add some more to the answer.

And so the first is the methodology development has not happened just in the isolation of the April file. When we started this work, we actually were initially working with the December 2014 file. And so the models that we've developed, the code that we've written, the way we've run it, differences we've looked for and tried to see what happens over time. We've done with December, we've done it with the April, which is what you see as part of the hospital dry run. And we're currently doing and have done most of with the July data set as well.

And so I think that that provides us some reassurance about it. I think the other aspects about the specific data issue that occurred in April, I'll leave to CMS. Kristie?

Kristie Baus: Hi, this is Kristie. We are very aware of the glitch that occurred with the April data. And we've put some quality assurances in place to make sure that that doesn't happen again. So when — if we do decide to move forward with the project, critical access hospitals will be receiving another preview report which will have all of the measures that they submit incorporated into their star ratings.

Aryeh Langer: Well thank you very much.

Operator: Your next question, you do have a followup from the line of Amy Chin.

Ismail Sirtalan: Hi, this is Ismail Sirtalan from Greater New York Hospital Association. I had a question about the region — regionality of these results of the ability of these results at a region level. And the star rating, as you have mentioned earlier, is a consumer-oriented measure.

So if I'm looking at this from a consumer's perspective, and if I want to pick a good hospital, how would I know that which hospital is better in my region because these are all national rated — nationally rated star ratings, and then there are 17 hospitals that are five stars. And I'm probably unlikely going to be able to go into one of those in my region. They don't exist in my region. So if I want to pick a good hospital in my region, and if there's little variation at the region level, how do you account for that?

Dr. Arjun Venkatesh: Thanks for that question. So we haven't really done many analyses looking about – looking at the regional distribution of star ratings. But I think your point is very well taken regarding the distribution which, you know, we acknowledge there's many three star hospitals as a result of the current kind of distribution of hospital performance on the existing measures on Hospital Compare.

And so I think that if there's specific comments you have around how this, you know, regional benchmarking could be done or ideas in that space, we're certainly open to that. And we encourage you to submit them to either the public comment process or through the Dry Run box, either one, for us to think about those and take those to CMS.

Ismail Sirtalan: Thank you.

Operator: As a reminder, to ask a question press star 1.

Your next question comes from the line of Chris Tomac.

Christopher Tomac: Hi this is Christopher Tomac with Sharp HealthCare. I heard a couple of references to if this program is determined to go forward along with no clear timeline being given. I was wondering what factors CMS is considering in making that final determination of if and when it goes forward? Thank you.

Kristie Baus: So currently what we're — we're looking at all the public comments that come in on the methodology report. And, you know, based on those methodology questions and comments, we will make a determination whether or not this methodology should, in fact, move forward.

Aryeh Langer: Thank you very much.

Operator: Your next question comes from the line of Alberto Luque.

Alberto Luque: Hi, this is Alberto Luque from Sylvester Comprehensive Cancer Center Miami. And it is somewhat related to this call. We know we are not eligible for the quality star rating. However, it is still not clear if we are eligible for the strategic – for HCAHPS Star Rating program. So, hopefully, you have the answer for it.

Bill Lehrman: Hi, this is Bill Lehrman at CMS. If your hospital participates in the Hospital Inpatient Quality Reporting and has more than 100 completed surveys for the four- quarter rollup, then we would compute the HCAHPS Star Ratings and publicly report on Hospital Compare.

Alberto Luque: All right, thank you.

Operator: And there are currently no further questions at this time.

Additional Information

Aryeh Langer: Well, we must have done a great job to end 10 minutes early. I just want to remind everybody that an audio recording and written transcript of today's call will be posted to the MLN Connects Call website. We will release an announcement in the MLN Connects Provider eNews when these are available, if you have any further questions, or you come up with anything in the next couple of days, we can refer you to slide 37. You'll find the resources there for further help.

On slide 39 of the presentation you'll find information and the URL to evaluate your experience with today's call. Evaluations are anonymous, confidential, and voluntary. We hope you will take a few moments to evaluate your MLN Connects Call experience.

Again my name is Aryeh Langer. I'd like to thank our presenters and also thank all of you for participating in today's MLN Connects Call. Have a great day everyone.

Operator: This concludes today's call. Presenters please hold.

-END-


CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES


Medicare
Learning
Network
Official Information Health Care
Professionals Can Trust


MLN Connects