

**ESRD Quality Measure Development, Maintenance, and Support Project**  
**End Stage Renal Disease Dialysis Facility Compare (DFC) Star Ratings**  
**Star Rating Post-TEP Conference Call #2 Minutes**  
December 2, 2015 1:00pm – 2:00pm (ET)

TEP Members	UM-KECC	CMS
Paul Conway	Yi Li	Joel Andress
Jane Pendergast	Joseph Messana	Elena Balovlenkov
Catherine Sugar	Richard Hirth	Sofia Martinez
Richard Cook	Claudia Dahlerus	
J. Richard Landis	Chris Harvey	
Franklin Maddux	Zezhi (Zac) Zhang	
Allen Nissenson	Natalie Scholz	<b>NORC</b>
John Reynolds	Cindy Liao	Rebecca Catterson
Tonya Saffer	Jennifer Sardone	
Chris Sarfaty	Jordan Affholter	
Nicole Stankus	Casey Parrotte	
David White		
Anonymous Patient A		

## Introduction

UM-KECC welcomed everyone to the Post-TEP Star Rating conference call, and thanked the TEP members for their time. Discussion for this call was to address TEP feedback and concerns about the star rating methodology.

## Updates on Patient Reported Outcomes

CMS informed TEP members that CMS is in the process of adding the ICH CAHPS (Patient Satisfaction Survey) measure to the DFC website for October 2016. CMS has also tasked UM-KECC to write a white paper that investigates available data and potential measures on patient reported and patient centered outcomes. The paper would include an evaluation of the kidney disease quality of life (KDQoL) data and measure. CMS is also investigating measures assessing infection (safety) and patient grievances. CMS stated they have received concerns about the usability use of the website and are working on simplifying the website. CMS is collaborating with the patient organizations, American Association of Kidney Patients (AAKP) and National Kidney Foundation (NKF) to develop consumer testing groups.

## Presentation Overview and Objectives (UM-KECC)

UM-KECC stated that the second teleconference on December 4, 2015 was scheduled to allow more time for TEP discussions for any topics that were not covered on this December 2nd call. Additionally a public comment period would be set aside from 1:55pm to 2:00pm on both calls.

UM-KECC briefly gave an overview of the agenda, which included these three topics: 1) the update for Patient Reported Outcomes on DFC, 2) options for updates to the methodology, and 3) feedback from the TEP on methodology options.

UM-KECC provided a context for the call objectives, explaining that they would be presenting a simpler scoring paradigm that uses absolute thresholds for individual measure scoring. This was an alternative to the current methodology based on relative performance and fixed percentages of star ratings for each year. The current methodology represents one end of a spectrum of methodological possibilities for the calculation of star ratings. What was being presented on this week's call represented the other end of the range of methodology options. UM-KECC stated the intent was that this would serve as a starting point to stimulate discussion among the TEP members. The purpose was to effectively broaden the TEP's discussion about available methodology options while exploring the advantages and limitations of each to allow UM-KECC to better identify areas where there is TEP consensus on a methodology.

### **Options for Methodology Updates (UM-KECC)**

UM-KECC reviewed four questions that defined the objectives for the presentation:

- How to set thresholds for individual measures?
- How many thresholds?
- How to translate measures with thresholds into a star rating?
- Should thresholds be updated over time?

UM-KECC outlined the first methodological question for the absolute threshold methodology, providing an example of scoring addressing suggestions and feedback from TEP members from the previous call.

Using the fistula measure as an example, UM-KECC presented a model of measure scoring using thresholds, where facilities that scored above a threshold would receive the full score of five points. While facilities that scored below the threshold would receive a score of one point. UM-KECC also stated more thresholds could be used for scoring an individual measure. This means a facility could receive a score between one through five points. Measures could also be scored continuously which would result in many thresholds representing a continuous range of measure scores (e.g., 0-100).

For this methodology presented, the final score would be calculated as the average of the seven DFC measure scores. The star rating cutoffs could then be determined by rounding the final score to the nearest integer (e.g., 4; 4.5, etc.).

UM-KECC furthered the example of absolute threshold scoring and presented analysis results using a low and middle threshold (low=55% and middle=65%) on the fistula measure, and then how the final star rating might differ using either threshold for all the individual measures. Using the lower threshold, more facilities achieved a full score, while using a middle threshold resulted in fewer facilities achieving the threshold. Over 900 facilities would receive a four and five star rating for passing the low thresholds, but receive only a one or two star rating applying the middle thresholds. Also small facilities may fail to pass low thresholds more often by chance than as a reflection of their actual performance.

UM-KECC stated that the analyses demonstrated that the star ratings are highly dependent on where the thresholds (or targets) are set for individual measures, suggesting that the choice of a single target may be arbitrary.

### **TEP Discussion and Feedback**

The Public Reporting/Patient and Consumer Understanding Workgroup chair commended UM-KECC and CMS for being responsive to the TEP concerns brought up during the previous teleconference call. The TEP chair stated the importance of showing options for a methodology that aimed at being simpler for patients and

consumers to understand. The TEP chair explained that information of great scientific importance is not mutually exclusive from information of great importance to patients. The TEP chair reinforced the importance of the methodology being in service to what gets measured.

The Methodology Workgroup co-chairs brought up several important statistical considerations when using scoring using thresholds. They stated the star ratings are sensitive to where the individual measure thresholds are set, noting it can be problematic if a threshold is set in the middle of a measure distribution because that would result in very different scores for facilities that may actually share little difference between their performance. It was stated that establishing one objective threshold would be difficult because it would require external expert opinion and clinical information on what is an appropriate and meaningful target. It was also stated that multiple thresholds would be worth considering.

One TEP member stated that scoring based on one threshold (also called binary scoring because it is pass/fail) may be an option for the highly skewed measures where there is much less difference in underlying quality because most facilities have achieved what represents good performance. In contrast, measures with a wider distribution may not be suited for scoring based on one or two thresholds because there is greater variation in facility performance on the measure. Another TEP member stated that a continuous scoring system would be preferred for measures with a wide distribution while using absolute threshold scoring could be applied only to highly skewed measures, or those measures could be considered for retirement.

The Public Reporting/Patient and Consumer Understanding Workgroup chair communicated the importance of getting measures on the website that are meaningful to patients such as cleanliness, infection, fistula use, facility practices in their referral for different modalities and renal replacement therapies (i.e. home dialysis, PD, transplant).

A few TEP members from the Public Reporting/Patient and Consumer Understanding Workgroup stated that the simple absolute threshold methodology presented and the current star rating methodology were both too complicated for patients to understand. What was most important was accuracy of the rating, and to make information on the methodology and measures simple while still making available layers of additional technical detail and information for those (patients) who want to know these details. Several TEP members reiterated the importance of how the website is presented and the interpretability of the website were also very important, such as improving the usability of the website.

With respect to scoring highly skewed measures, one TEP member noted the importance of still measuring performance of the facilities in the tails of the distribution (outliers with very poor performance on the measure). They stated that the tails of the measure's distribution may show which facilities have bad quality care and which have superior quality care (outliers with very good performance).

The Methodology Workgroup co-chair explained that thresholds would be absolute if clinical evidence showed where care is good or bad. The TEP co-chair explained that care in the middle of the distribution could be really good, but without clinical evidence, selecting an absolute threshold would be relative. The TEP co-chair further clarified that "absolute means someone sets an external reference for what is an adequate score, and relative means we based it on their positions and how far away they are from each other without regard to whether we actually think that's clinically adequate or not."

The Methodology Workgroup co-chair summarized the two main issues discussed during the TEP call. The first issue, regarding methodology, was how to get an accurate score. The second issue was how to present scores in a clear and easy to interpret way. The Methodology Workgroup co-chair stated the methodology can be flexible to the objectives of public reporting that need to be accomplished. The chair stated that even if the

methodology is complex, that does not preclude it from being explained in language that consumers can understand and interpret.

The Public Reporting/Patient and Consumer Understanding Workgroup chair concluded by restating the importance of having measures that are meaningful to patients, and that the website needs to be effective in providing consumers with information that helps them make informed decisions about the quality of dialysis care at facilities, since this care will have a direct impact on their lives.

### **Public Comments**

The following public comment was transcribed verbatim from the teleconference recording.

Jackson Williams (Dialysis Patient Citizens): “So I just want to respond to what was said. I think the normative question of what constitutes a passing score or what ought to constitute a passing score is tough, but I really think the best inquiry is what is consumers’ expectations of what these symbols mean. To the extent of the final distribution is close to the distributions of other star ratings, whether it is what people are familiar with from Yelp or from the other CMS compare websites, or the other federal websites, like say for Cars.gov. I think that could provide a lot of good guidance for what ultimately the distribution of the stars would look like. Thank you.”

### **Closing Remarks**

UM-KECC provided some closing remarks for the TEP meeting including the importance of TEP discussion surrounding the question of how to accurately measure quality of care.

UM-KECC stated that the TEP discussion will continue on the follow-up TEP teleconference call scheduled for December 4, 2015. UM-KECC thanked everyone for participating in the teleconference call, and for contributing to the discussion.