

ESRD Quality Measure Development, Maintenance, and Support Project
End Stage Renal Disease Dialysis Facility Compare (DFC) Star Ratings
Star Rating Post-TEP Conference Call #3 Minutes
December 4, 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
J. Richard Landis	Claudia Dahlerus	
Allen Nissenson	Chris Harvey	
John Reynolds	Zezhi (Zac) Zhang	NORC
Tonya Saffer	Natalie Scholz	Patti Truant
Chris Sarfaty	Cindy Liao	Rebecca Catterson
Nicole Stankus	Jennifer Sardone	
Sumi Sun	Jordan Affholter	
David White	Casey Parrotte	
	Karen Wisniewski	

Introduction

UM-KECC welcomed everyone to the Post-TEP Star Rating conference call, and thanked the TEP members for their time. UM-KECC summarized three key points from the previous TEP call that occurred on December 2, 2015.

1. First, there is a critical need to include measures relevant to patients and consumers on the DFC website and in the star ratings in order to provide meaningful information to these consumers. In addition, the usability of the DFC site and star rating system are critically dependent on how these data and information are presented in order to be interpretable by users.
2. The second point, regarding the use of absolute threshold/targets and several points made by TEP members is that 1) setting thresholds/targets for individual measures would not be a simple undertaking and would require substantial time and effort in order to define appropriate targets, and (2) the use of multiple thresholds was generally preferable to using one threshold, particularly for those metrics with continuous distributions.
3. The final summary point is that multiple TEP members expressed opinions that use of a specific statistical method was of less importance, but what was very important was to use the statistical methods that would result in the most accurate information about facility performance, and, that what is presented should be straight forward for patients and consumers to understand. The Public Reporting/Patient and Consumer Understanding Workgroup chair summed up the discussion best with the closing question “can someone (sic consumer) take a look at the website and have confidence in making a decision that impacts ultimately, the quality of care that can either extend and prolong a life or jeopardize the quality of or the life itself?”.

UM-KECC asked the TEP to confirm that the summary points presented were an accurate reflection of the December 2, 2015 TEP call. The TEP members, including the TEP chairs, confirmed that the summary points were an accurate reflection of the discussion from that call.

UM-KECC stated that the other critical topic where TEP input is needed relates to how the individual measure scores are translated into star ratings. The current star rating method provides information about how each facility performs on the DFC measures relative to other facilities in the current reporting period. Many in the dialysis community have commented on their preference for a system that rates each facility on how it performs relative to some baseline standard, defined either by peer group performance in a baseline year, or by defining absolute thresholds for individual measures. UM-KECC's presentation highlighted areas of discussion where TEP feedback was needed on these issues.

Threshold Presentation

On the previous teleconference call, UM-KECC presented options on the number of measure scoring thresholds (targets) and placement of these scoring thresholds. UM-KECC stated that facility performance on measures will likely change over time. By setting a threshold, it is implied that the target will remain fixed for some period of time. Facilities will then be scored each year on whether they obtain that fixed target.

UM-KECC discussed how improvement on measures may change measure scores and in turn the distribution of star ratings. Overall improvement on the measures by facilities would result in higher star ratings over time (upward drift). This differs from the current method which does not allow upward or downward shifts in the percentage of facilities receiving for example, five stars, or one star. Over time if measures were removed, because they were skewed the star rating distribution would likely shift back to something more close to the original distribution. To illustrate this, UM-KECC presented results of a simulation showing the impact of change over time.

UM-KECC stated that it may therefore be necessary to consider how to set or reset the measure thresholds and/or baseline year when measures are removed or added to the star rating.

For the TEP discussion about the baseline year, UM-KECC provided three examples for how to set measure thresholds. The first option would be to set fixed measure thresholds in an initial or baseline year and set fixed star rating cutoffs based on the distribution of the final combined measure scores in that year. For this option, a rule may need to be created for whether the thresholds would need to be updated to a new baseline year every time measures are added or removed from the star rating. Another consideration is whether a period of time and additional criteria need to be pre-specified for what would trigger a re-setting of measure thresholds.

The second option briefly discussed is the current methodology which updates measure thresholds relative to the current year distribution of measure scores. This can be seen as re-baselining the measure thresholds every year.

The third option is a version of setting the baseline year to fixed measure thresholds as well as using relative thresholds (thresholds relative to national average each year). For example, measure thresholds for the performance year would be determined two years ahead of time based on current year's data. For example, thresholds for 2017 would be based on the distribution of 2015 data. This method would allow facilities to better predict the scores they need to attain (e.g., for a 5-star rating) while minimizing the consequences that might occur due to large measure drift or as a result of adding or removing measures. The thresholds would change each year as determined by moving forward one year (i.e., 2016 data would then be used to set thresholds for 2018; 2017 data for 2019, and so forth).

TEP Discussion and Feedback

The Methodology Workgroup co-chair clarified that picking a baseline year still is a definition of a relative threshold, because it is a relative measurement from the baseline year to where the scores are now. The co-chair stated that, ideally, absolute thresholds should be based on clinical opinion about what constitutes quality clinical outcomes, but absent no clear clinical consensus on what is good clinical care, thresholds will ultimately end up being relative, i.e., facility performance relative to others versus an externally defined standard. If an absolute standard is used then shifts in scores to more four and five stars for example, should only represent actual change in the quality of care, otherwise changes may represent artificially inflating measure performance.

The Methodology Workgroup co-chair also cautioned against removing highly skewed measures as one would still want to identify very poor performance for those facilities in the tails. The co-chair stated that binary scoring may be an option for such skewed measures, but that how many thresholds to use for a measure will depend on each measure's distribution. For the measures that have a more continuous distribution, binary scoring would not be a good option as it will result in facilities that are very similar getting very different scores (that is, differentiating quality when there are no real differences in quality because nearly all facilities are achieving very good performance). With these points made, it was stated that it is important to have a practical and relatively simple system instead of reviewing every possible consideration in the pursuit of a perfect system.

The Public Reporting/Patient and Consumer Understanding Workgroup chair agreed with the importance of moving forward and reinforced the importance of the addition of measures that are meaningful to patients (patient reported outcomes), making the website easy to use, and making sure to use the most appropriate methodology, deferring to the experts on the most appropriate statistical methods that accomplish the goal of providing usable and easy to understand information to consumers, where methodology functions in service to policy. The chair confirmed that patients want to know the quality of their care especially about facilities near them, and that patients may also be willing to drive farther to go to a facility that scores better on a particular measure.

One TEP member stated that the website has a lot of information and can be very helpful. They would also like to see other information added to the DFC website, such as state violations (of the Conditions for Coverage ESRD Regulations), or clinic complaints, and whether facilities accept transient patients for appointments. The latter they said would be helpful because transient patients may rely on the DFC site when traveling. Two TEP members went on to express concern that poor clinical outcomes, for example on the Kt/V measure, may be related to patient non-compliance, and not be under the complete control of the clinic, therefore the measure may not be a complete reflection of facility quality.

Other comments about how to set the baseline year included a concern from one TEP member that shifts over time in thresholds should not result in fewer facilities with four or five stars. The TEP member stated that the standard by which four and five stars are based should remain at the same threshold or improve over time, which would result in more four or five star facilities. When asked, they said they do not feel strongly about which measures are scored using absolute thresholds but that what is important is that the description of measure scoring is clear and transparent, so that, for example, patients will understand the comparison of facilities is to the national mean.

One of the Methodology Workgroup co-chairs asked the Public Reporting/Patient and Consumer Understanding Workgroup members whether for skewed measures, if patients would be most interested in knowing what are the outlier facilities in terms of performance, that is, those that are in the tail that are trailing far behind all other facilities in terms of performance. Several members of the Public Reporting/Patient and Consumer Understanding Workgroup answered yes, and agreed that the main information that is important to patients is if

performance in the tails (very poor scores) signal lower quality or a potential harm to patients. Additionally, one member stated that distinguishing superior performance is also of interest, in order to identify facilities that are not only passing the “pass/fail” threshold but are far exceeding the threshold.

Based on this discussion the Methodology Workgroup co-chair summarized that for skewed measures it may make sense to establish one threshold.

The discussion then moved to a second question posed by one of the Methodology Workgroup co-chairs. They asked members of the Public Reporting/Patient and Consumer Understanding Workgroup, for measures that have a wide distribution, is it important for patients to distinguish the measure scores based on multiple thresholds distinguishing performance. The co-chair used the analogy of the standard school grading system, to distinguish which facilities may receive a middle grade like a B, C, versus a D or F grade. The fistula measure was used as an example of such a measure because it has a fairly spread-out symmetrical distribution.

The Public Reporting/Patient and Consumer Understanding Workgroup chair stated it was important to show a spread on the fistula measure, and stated that three categories (fail, pass, and high pass) would be appropriate for that measure. Two TEP members concurred they would like to see more gradations/categories of grades on some measures (like fistula) because differences in performance could be impacted by patient behavior (non-compliance) therefore for them the issue was whether several levels was enough or does the system need to include more categories for scoring a facility on these types of measures. Two of the TEP members agreed that the number of gradations/categories would also depend on the measure.

One TEP member summarized three distinct areas of the TEP discussion about thresholds and quality measures. First, discussion has focused on what measures should be included in the DFC star rating. Second, some of the additional information patients are interested are not measures of quality but related to informational and logistical information about facilities. For example, modalities offered is more logistical information for patients, not a measure that determines quality. Third, with respect to the scoring methodology, the TEP member agreed with the Methodology Workgroup co-chairs that binary scoring (one threshold) would be most appropriate for skewed measures like Kt/V, and that other measures with a wider spread would not be a fit for such thresholds but instead would could be scored continuously, or using a mixture of methods (as described earlier in the presentation). The TEP member further stated that it may be difficult for a group to establish consensus on absolute thresholds for a measure because getting expert consensus on what is good or bad quality is not really feasible because there is not enough scientific evidence to determine those standards.

The Methodology Workgroup co-chair agreed that continuous scoring or a mixture of methods would be ideal over purely absolute thresholds for individual measures that are not skewed, because it would provide patients with more relevant information.

TEP members were again asked to share any additional thoughts on the threshold and baseline year options. One TEP member reinforced the importance of adding measures of patient experience of care, staff responsiveness, and staff level of education in dialysis, which was discussed at the in-person TEP.

In summarizing the discussion of whether to fix thresholds, one of the methodology co-chairs stated there are two potential options depending on whether patients want to be able to see year to year improvement. One option would be to fix an absolute threshold in the baseline year and the rating each year would then be relative to that baseline year threshold. This would allow one to see improvement of facility performance over time, from the baseline year. The other option is that each year the threshold moves and is relative to performance in that current year. They suggested this would be the preferred method. In sum, fixing to a baseline year shows more of how things have changed versus performance relative to the current year. They explained that how

thresholds are set depends on the type of question one wants answered about facility performance: to see improvement over time or, to see how facilities perform relative to one other each year.

Public Comments

The following comments are transcribed verbatim from the teleconference recording.

Kathy Lester: “Well, thank you all for the opportunity, on behalf of Kidney Care Partners. You have seen our letters so you probably know about what I am going to say. Joe [Dr. Joseph Messana] asked the question about the current five star methodology versus benchmarks or absolutes. And as you know, the patient organization KCP and others in the kidney care community believe that quality is not a relative concept. And while we do understand the differences between some measures being able to have an absolute threshold, for example the Kt/V, while others like the standardized ratios may not have one, we think that there is a way to provide the absolute quality that is being provided. And we recommend not that you have to adopt QIP methodology, but to look at that as a way -and you can tell the difference- between five star results and QIP results, by just looking at the five point scale that the QIP has versus the even buckets of star distributions that currently exist. We would encourage you to do that as a way to think how you might have that mixture of methods.”

Hrant Jamgochian: “This is Hrant with Dialysis Patient Citizens. I just wanted to first actually commend not only just this discussion, the previous discussion, the flexibility and openness with looking at what other potential alternatives and options there may be to make it more meaningful and useful for patients. And in particular, the one comment I would like to kind of add -I know it was just made before the comment period opened up- is the importance of the patient experience. It’s the number one response that our members gave us on what matters to them is “does my facility listen and treat me with respect?” That is the kind of quality of care, and there isn’t a specific particular measure for that. Obviously there are multiple questions from CAHPS that could be used to get to that kind of an issue, but those kinds of issues I think are the kind we would love to see CMS ultimately look at incorporating into the star ratings. And the one caution I kind of would like to encourage the TEP to kind of keep in mind or take into consideration is -absolutely I was really glad to see that certain measures made sense to have a pass or fail and others more gradation- but the one concern is I think there is a lot of substitution of the gradation, the “a-b-c-d-f,” with a “5-4-3-2-1” star. I think most star ratings there is just a real stigma if it is a 1 or 2 you don’t want to use it. So when you do the conversion of when you are looking to decide which measures or which metric, if it getting good quality of care, or if just right below average, does that mean that it’s a two star and you shouldn’t go to that facility- if that is something that is important? I saw some of potential distributions. I think if someone deserves a two star, I think our organization and patients have said, then give them a two star. But to have an arbitrary we’re going to put 10-20-30 percent in these various different buckets, that just I think is kind of concerning. I appreciate the discussion about “should there be an absolute?”, “should there be a relative measure?”, and how to get there and really encouraged. Even though it might be easier to fall back on well we can’t get it perfect, so let’s just get something- I want to strongly encourage the TEP panel to continue this work and how much time it is spending on this effort because the closer you can get to perfect -I know it’s not going to be perfect- please keep trying to get as close as you can so making it as meaningful for patients. Thank you for everything you are doing.”

Closing Remarks

UM-KECC thanked everyone for participating in the teleconference call, and for providing their input to the TEP discussion. All TEP members were encouraged to email UM-KECC with additional comments regarding 1) the TEP teleconference discussion or 2) any comments specifically regarding scoring using a system that measures relative to other facilities in the current reporting period (similar to the current system) vs. a scoring system that fixes thresholds to a baseline year.

Appended TEP emailed comments after the TEP Teleconference Call

Tonya Saffer: “I agree with the patient TEP members that it’s hard to identify the methodology when the measures could (and should) change to be more responsive to what patients value as quality care. I think everyone agrees we should have absolutes where we have consensus and benchmarks that perhaps rely on national averages in their absence. Given the current measures on DFC, the methodology that may be in most service to those measures is similar to what is used for the QIP because it assigns points for meeting ranges and those points perhaps could be translated into stars. This seemingly would help to address the issue where there is a lack of differential performance on certain measures – I would love to hear what Jane and Catherine think about this. However, what is lacking with the QIP is the fact the data is two years old by the time it’s reported. From my discussions and interactions with patients and my own perspective as a daughter-in-law to a stage 4 CKD patient, the performance rating should reflect the most current data in order to be most relevant to patients decision making. What might be interesting to be able to see in DFC is also historic performance on the stars, but this will be more challenging if measure sets are constantly changing.”

John Reynolds: “Only sitting for 4 hours with your blood filtering through a machine can you really get a good idea of what quality care means to you. I really do not care what system is used plotting points on a graph or a sliding scale to infinity. Strictly, as a dialysis citizen, I cannot find a lot of fault with the DFC web site. If in the future, either through Medicare surveys, state inspection filings, or complaint logs, there is a place for a more transparent picture of dialysis facilities then that will only be a plus for consumer information. Thanks to everyone again for allowing me the opportunity to take part in the TEP process. My only regret is that I lacked the eloquence and intelligence of so many of my fellow participants.”

David White: “I'd like to thank everyone for assisting me in getting a better grasp of the methodological issues that we are discussing. I would prefer a baseline year scoring system because it would allow tracking a facility’s improvement or decline over time, both in its case and relative to other facilities. The QIP addresses this to a great extent, but the data might be useful in the future. If this premise is correct, could said improvement or decline over time be captured if the performance threshold was adjusted annually?”

In response and addition to David White’s comment, Dick Landis added the following comment.

Dick Landis: “I agree completely. A recent year benchmark would allow absolute quality comparisons, as well as relative quality comparisons to be utilized in weighing these comparisons. I would also favor these absolute and relative comparisons being available for primary sub-domains, in addition to the overall summary 5-star rating. The analogy would be a restaurant w/ a 4-star rating overall, but has a 5-star rating on quality of food, offset by only a 3-star rating on the service. This would allow patients and their families to understand the primary components influencing the overall rating.”

Chris Sarfaty: “Thank you for the opportunity to review the minutes of our last 2 TEP calls and thank you for the opportunity to submit comments. I recognize these comments are a bit long and at times a bit broad, but I do think they are relevant. They have been percolating in my mind as I have observed and participated in our meetings and conference calls.

Like many other TEP **public reporting/patient and consumer understanding group** members have stated, I too, am thankful for the attention that is being paid to the effort of incorporating concerns and questions from the patient/consumer point of view about the 5 Star Dialysis Rating system. While the process may be hard work, it is important work, and we are learning things every step of the way. Perhaps what we learn as we struggle through our process will be helpful to other programs.

I want to frame my comments by mentioning my point of view and mentioning the TEP charter.

First, my point of view. I worked in dialysis directly with patients for 25 years, and I also worked directly with social work staff mainly, but also with all members of staff, to support them in their efforts to provide patient-centered care. My point of view is also informed by my experience as child watching our close family friends perform dialysis in their living room for hours on end with a machine the size of a small foreign car. I also watched and try to help my father and step-mother as my father fell into kidney failure due to diabetes. Living in a rural area, and preferring to be at home, he first attempted peritoneal dialysis and eventually moved to hemodialysis when his residual kidney function (which helped him to sustain peritoneal dialysis) was lost secondary to a dye reaction during the process of repeated attempts to create a fistula. Despite the master's level education of both of my parents, the expertise required to manage the many doctor's visits, the many medications, run the home dialysis, attend in-center dialysis and follow the dietary and fluid requirements was very stressful for them. In addition, they would have never been able to navigate the DFC website. They were already a-wash in hand-outs and booklets and med lists and calendars. There are so many decision-tree forks for the CMS webpages, to get to the desired information, they would have been lost, given their low computer literacy level.

Secondly, I want to mention the charter. The Patient Protection and Affordable Care Act from the federal government calls for information to be transparent, easily understood, and widely available, so as to improve quality and satisfaction. I interpret that to mean that the clinical values should be communicated in a way that the lay person can come to understand them and that content should include information that patients want to know as consumers, including, especially, patient reported experience of care.

My comments after the December 2015 conf calls:

Patients have varying needs for a central location that will give them helpful data and information about dialysis centers. In our TEP meetings, our discussions have come to refer to **"choice"**; patients choosing which dialysis center to attend, however "choice" may not be the main use of the 5 star rating. Based on my 25 years working directly with patients on dialysis, based on my experience of my father on dialysis, and based on comments made by many of the consumers of dialysis in the **public reporting/patient and consumer understanding group** during the 2 day Baltimore TEP meeting, often there is not, in reality, a lot of choice involved. People often end up in dialysis centers based on where their doctor rounds, based on where their insurance will cover them, and sometimes based on transportation options available. In addition, it was noted that many times, especially when a patient has not had pre-esrd care, patients are often too ill or too overwhelmed to even begin to research dialysis center choices when they are first beginning dialysis.

When dialysis consumers travel, that is the situation in which we more often see the patient needing to make a choice about where to dialyze. The number of patients who travel, relative to the whole dialysis population is relatively small.

There are benefits, in addition to "choice", of a central location that will give helpful data and information about dialysis centers; that is the benefit of both patients and staff having **the opportunity to learn more about their center**. This provides an opportunity for centers (staff and patients alike) to learn from other centers; what goes well, what is a challenge, and possibly, especially within companies, to share methods for improvement. Not a

shame or blame paradigm but a rising-tide-lifts-all-boats paradigm. After all, we want all centers to do well, because we want all patients to be as well as possible. Learning about your center once you are there, being empowered by knowledge from the website, in addition to knowledge from other sources, gives you a chance to then know how to help your center get better and what your rights are to participate in that improvement process.

Unlike hotel chain rating systems or restaurant rating systems where you have a meal or spend a night and then maybe don't return again, in dialysis centers both patients and staff do return again and again. In the case of patients, quality of life and longevity are what is on the table, not a meal or an overnight bed. In the case of staff, quality of work life and seeing patients do well and thrive is at stake, not whether a hotel or restaurant customer you may never see again is happy with your service or not. Though there surely are times when staff and patients have difficulty communicating with each other, as chronic dialysis can be arduous for both staff and patients, both want to leave their center at the end of the day feeling good about how things went, especially since both have to return and work with each other again. The bottom line; patients want to be able to do well - staff wants patients to do well.

So, unlike a hotel or restaurant rating for a one time experience that tells you "yes" this is a good place to eat or "no" don't bother to sleep there, our dialysis center ratings can be seen as ***sharing information*** for the common good; the common learning and on-going improvement that should be inherent in any medical setting. Improvement is inherently a continuous process.

With the health and well-being of patients in mind, and the ***many things*** that contribute to that health and well-being, it seems near impossible to come up with a single star rating.

In Baltimore, in the **public reporting/patient and consumer understanding group**, there was some discussion about the idea of a website that would allow you to select from a menu of items, the information about a dialysis center that is important to you. In this hypothetical scenario, each item would be rated, and you could see where a center stood on a scale from low to high.

This list of items could include both the clinical outcomes measures as well as patient experience of care measures. And, ***patients do want to know about clinical measures, but they also want to know about experience of care measures***. They are both important to patients. Once it was decided how the rating would be determined, centers could be compared to each other as well; preferably all within as current a time frame as possible. In other words, as Tonya Saffer stated in her comments, not using data that is from several quarters ago.

I don't know how to weigh in on the discussion about relative vs absolute ratings. I don't know enough about the statistical implications and requirements, however, what I do know is that whatever choice is made, it will need to be explained very clearly to the public, dialysis consumers, public, and providers, and dialysis staff alike. This will help avoid misunderstanding and distrust of the website and our TEP process. Careful attention will need to be paid to how to educate, get the word out, and make the explanation available. In Baltimore, in **public reporting/patient and consumer understanding group**, Paul Conway shared some well-developed strategies about how to get a message delivered to the public using various media modalities.

Combining Statistical perspectives with public consumer perspectives is a challenging task, at the very least.

I am reminded of something Jane Pendergast's spoke about in her comments at the end of day 2 of our meeting in Baltimore. This may not be a direct quote, but I think its close. She said ***"if your question is not clear, the methods approach will not be clear"***.

I do think that in some respects our "question" is not clear. It is not clear in that we are asking so very many questions. In addition, it seems that different questions match best to different scoring methods. Trying to reconcile all of that into one star seems to defy the odds.

Drawing from my training in looking at the whole picture or whole system, it seems to me, that the repetitive nature of our dialogue, about HOW to measure and HOW to "grade" dialysis units is telling us something. From Day one starting with the live TEP meeting in Baltimore, through all the conference calls, and including comments on the summary report we have struggled with how do we best measure dialysis centers and produce a star rating that is meaningful and useful to consumers.

Is part of the challenge related to trying to measure something too big to fit into a star rating scale that produces a single star?

Above I mentioned the idea that surfaced in the **public reporting/patient and consumer understanding group**, of selecting items of interest to the consumer from a menu. Below is a slightly different idea.

If we are trying to measure something too big to fit into a star rating scale that produces a single star does that mean there should be more than one star? Perhaps a star rating scale for patient reported experience of care and a second star rating scale for clinical laboratory outcomes measures?

What would that look like?

Would one page be something like this:

This is what patients reported about how they experienced care at their center regarding:
for example (just a sample, not definitive):

- physical comfort
- center cleanliness
- staff* receptivity to questions
- staff* flexibility in helping to solve complex questions
- staff *friendliness
- staff* helpfulness
- patients feeling responded to a respectful manner
- staff* actively seeks patient feedback and response about their care
- easily available local transportation
- infection rates
- referred to modalities

(This could be based on patients who responded to an inquiry, i.e. Consumer Assessment of Healthcare Providers and Systems (CAHPS), or something similar?) (which would include a caveat clarifying that not all patients responded, and if possible, the percentage of patients from that clinic who responded.)

*Staff defined as all employees, from front desk secretary through physician and management

Would the other star rating page be something like this:

This is how facilities rated on clinically measured outcomes currently on 5 Star

- Anemia
- Dialysis Adequacy
- Vascular access
- Mineral and Bone Disorder
- Hospitalization and deaths

Thank you for this opportunity to provide comments.”