



March 3, 2015

Andrew Slavitt
Acting Administrator
Centers for Medicare and Medicaid Services
200 Independence Avenue, SW
Washington, DC 20201

Dear Acting Administrator Slavitt,

I am writing on behalf of Kidney Care Partners (KCP) to share with you our second set of consensus recommendations for the ESRD Five Star Technical Expert Panel (TEP) to consider. These recommendations focus on improving the structural methodology of ESRD Five Star to ensure that it provides accurate information that patients, their family members, and other consumers can use to make meaningful choices among dialysis facilities. These recommendations are in addition to the recommendations regarding standardized ratio measures that KCP provided to you on February 13, 2015, and the structural recommendations described in our letter dated October 30, 2014. KCP continues to strongly support providing patients, their family members, and other consumers with meaningful information to assist them in making health care decisions, including their choice of dialysis facility. ESRD Five Star should be significantly strengthened to achieve this goal.

Many of the recommendations below parallel those in our October letter, with one important difference. KCP strongly encourages CMS and the Five Star TEP to establish clear performance benchmarks for the non-standardized ratio measures. Because CMS has chosen not to apply benchmarks to define quality with regard to dialysis facilities, in sharp contrast to other CMS Five Star programs, the dialysis facility stars currently do not accurately represent the very thing patients, their family members, and other consumers expect – a measurement of actual quality performance of dialysis facilities.

Patients, other consumers, and providers do not view quality as a relative concept. The Robert Wood Johnson Foundation has highlighted how using stars to represent relative performance can mislead consumers:

Symbols can misrepresent performance if they show how providers scored relative to each other without indicating how the average performance compares to state or national benchmarks. If the overall average is low, a provider with five stars is not really providing good quality care. If the overall average is high, even a provider with just one star may be

providing excellent care. In the latter case, the symbols may be making a distinction that some providers will regard as unfair.¹

As currently constructed, ESRD Five Star does not accurately reflect facility performance, by focusing only on relative performance, at both the measure and overall score levels. The current methodology may force facilities to the bottom of the distribution, even when performance is objectively good and very similar to that of facilities with more stars. In his analysis of ESRD Five Star, Dr. Vince Mor, Professor of Medical Science, Florence Pirce Grant University Professor, Health Services, Policy & Practice, Brown University, describes the problems with the current methodology as well. “Transforming dialysis facility performance scores into ranks creates variation where there may be no meaningful variation and ignores the real information that currently exists in performance benchmarks.”² He further explains, “Creating relative rankings of providers can obscure the information people want to know by equating all the component measures into a composite that standardizes on an ambiguous average rank.”³

Patients and consumers, as well as providers, regard the current program as inaccurate and potentially misleading for the very individuals CMS seeks to empower. Defining quality of care using benchmarks is a necessary first step to addressing this fundamental problem in ESRD Five Star.

The Medicare Payment Advisory Commission (MedPAC) also has raised concerns about the current approach to ESRD Five Star.

The Commission believes the quality measurement process needs greater simplicity and clarity. Moving to two systems creates greater uncertainty. Furthermore, the Commission generally believes that the measurement of quality performance should be based on absolute standards rather than one calculated from the performance distribution. We believe this is fairer to providers and gives clearer targets for providers to meet. If CMS can establish the need for a Star Ratings System for dialysis facilities, the agency should then describe why it believes the measures in the QIP are an insufficient basis for establishing a Star Ratings Systems.⁴

Another related change that should be made is the elimination of the forced bell curve. KCP continues to strongly oppose the bell curve methodology. Despite

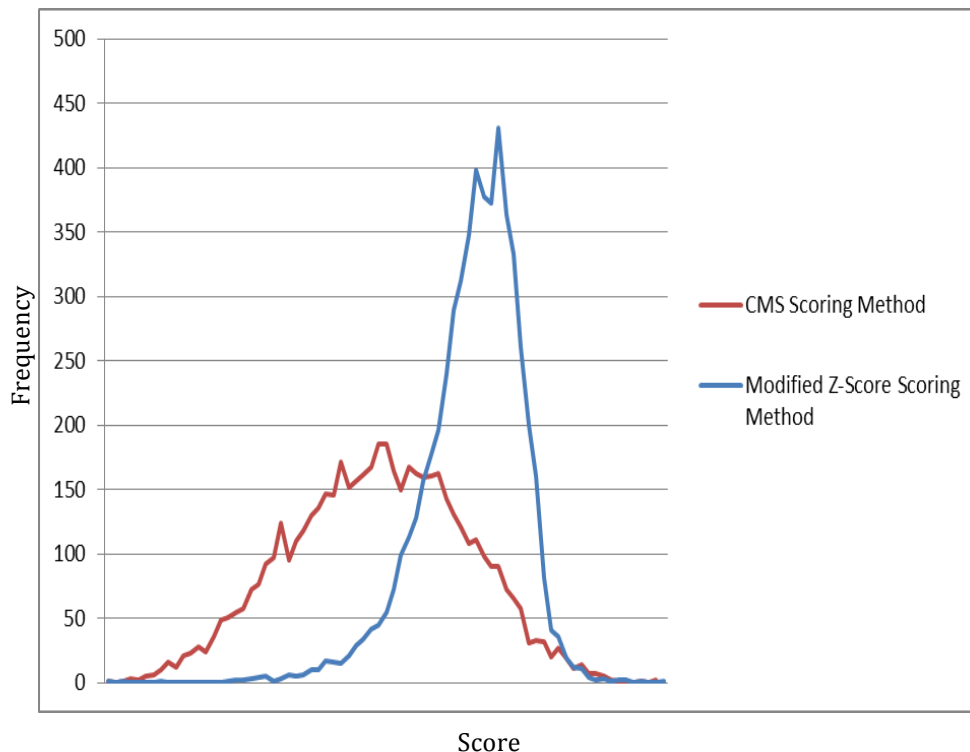
¹ Robert Wood Johnson Foundation, “How to Display Comparative Information That People Can Understand and Use” 17 (2010).

² Vince Mor, Ph.D., Memorandum to Kidney Care Partners “Comments on proposed ‘Five Star’ ranking system of dialysis facility quality” (Sept. 25, 2014).

³*Id.*

⁴MedPAC, Letter Administrator Tavenner (Aug. 15, 2014).

statements to the contrary, the methodology adopted by CMS forces both the measure scores and the overall star distribution into a bell curve. Such a model might make sense if underlying performance follows a bell curve pattern. In fact, the distribution of actual performance is asymmetric, with most facilities performing above the mean. This is illustrated by the graph below, in which a Z-score score model⁵ is compared to CMS's current percentile scoring model. The Z-score preserves actual differences in performance among providers. As you can see, the distribution is not a perfect bell curve. The difference in the two models is significant.



In addition, we reiterate our recommendations that CMS should:

- Score the measures using a Z-score model with performance benchmarks based on the 90th percentile of national performance;
- Use categorical results for the standardized ratio measures to account for the statistical uncertainty with numerical values;

⁵A Z-score is the number of standard deviation that is a given result is from a defined benchmark. In many applications, the Z-score is expressed as a number of standard deviations from the mean; for this program, we are recommending that the Z-score be expressed in relation to the 90th percentile of performance.

- Use star category boundaries that are based on actual variations in overall performance; and
- Set the star category boundaries to provide statistical confidence in the final stars.

KCP supports transparent quality programs that provide accurate and useful information to patients, their families, and other consumers. Unfortunately, the current CMS methodology used to determine facility-level stars requires significant improvement if it is to provide either.

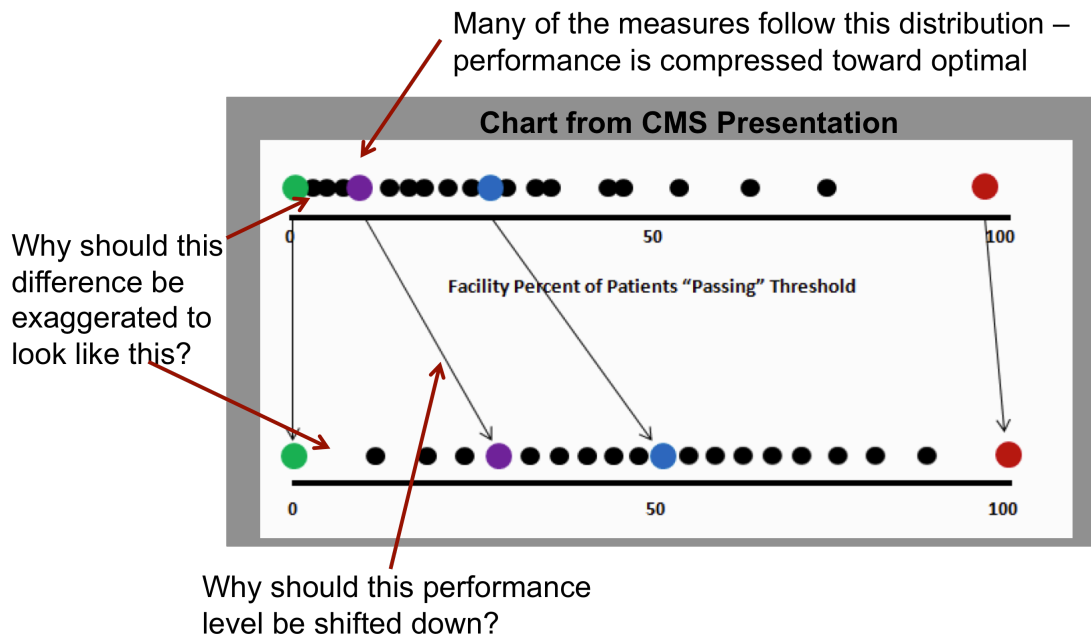
I. Recommendations for ESRD Five Star: Establishing an Appropriate Methodology for Determining Stars

As we have previously discussed, KCP and other members of the kidney care community have serious concerns about the methodology CMS has adopted for ESRD Five Star. The CMS methodology does not accurately reflect actual performance and in many cases does not represent statistically or clinically meaningful differences in quality. Thus, we strongly urge CMS and its ESRD Five Star TEP to adopt the following recommendations that would rectify the current shortcomings in the program.

A. Recommendation 1: CMS should score the measures using a Z-score model with performance benchmarks based on the 90th percentile of national performance.

We recommend that CMS combine the Z-score with a 90th percentile benchmark to score the measures. This approach would accurately depict real differences in facility performance, while setting the quality target at a high level. Unlike the percentile methodology, a Z-score methodology preserves the relative magnitude of performance differences across facilities. As the Agency's own data show, the performance distribution among facilities is not symmetrical. It becomes symmetrical only when CMS applies a methodology that distorts the actual performance distribution.

The current measure scoring model distorts differences in facility performance. The chart below comes from slides used during the initial CMS National Provider Call in which Agency staff described the ESRD Five Star methodology. The top line represents the actual performance of facilities, while the bottom line shows how, in applying the percentile methodology, performance is distorted.



The application of the percentile methodology distorts the difference among facilities, in some cases taking small differences and magnifying them, in other cases shrinking significant performance differences.

Z-scores would allow CMS to describe performance by expressing it as a number of standard deviations from the performance benchmark—a concept that is more readily understandable. To apply this methodology, CMS would calculate a Z-score for each facility for each measure. Then, CMS would average the scores within each domain. Finally, CMS would average the domains for an overall score. Using the most recent data available from CMS, the application of the Z-score methodology and the use of categorical values for the standardized ratio measures shows that the distribution of performance is not a normal bell curve, but rather an asymmetrical curve.

CMS should also adopt the 90th percentile of performance as the scoring benchmark for the non-standardized ratio measures. This approach would send the signal that high performance is the program goal. Unlike the current ranking model, which focuses on the relative position of facilities' performance, a true rating model focuses on actual performance relative to a benchmark.

Our proposed model is similar to that used for the Nursing Home Five Star staffing and quality components. That program utilizes fixed performance benchmarks instead of specifying set percentages for each star category for these components. As CMS has noted in the Technical Users' Guide for Nursing Home Five Star, the advantage of fixed cut points is that they allow the distribution of ratings to

change over time.⁶ Using the 90th percentile of performance as the benchmark for ESRD Five Star would achieve this goal, but also allow the benchmark to shift upward as overall performance improves.

The importance of relying upon benchmarks to define quality is evident from a review of the Kt/V dialysis adequacy measures. Dialysis facilities perform exceedingly well on this measure. About half of all facilities have 90 percent of their patients with a Kt/V greater than 1.2. The strong performance on this measure should be expected. Dialysis adequacy metrics measure the core component of care provided by dialysis facilities – how well they remove toxins from patients’ blood. In addition, CMS has been measuring dialysis adequacy in one way or another since the mid-1990s through its claims monitoring program. Yet, when forced into a ranking, facilities with the same level of performance may be several ranks apart. This distortion of actual performance misleads patients and consumers about the quality of care they are receiving.

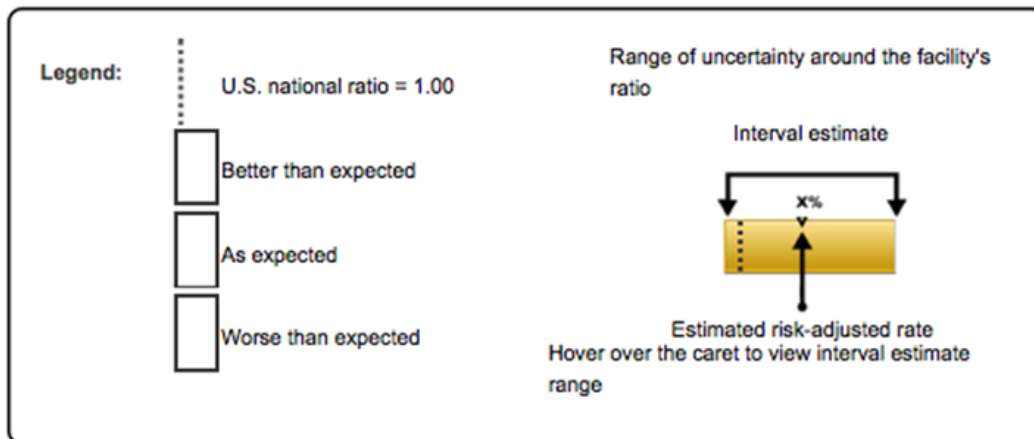
Using a 90th percentile benchmark, however, would provide useful information that is not misleading. By selecting this benchmark, CMS would be setting a benchmark that goes beyond simply comparing a facility with national averages. By doing so, the Agency would provide value for the continuous quality improvement process.

Thus, we strongly urge CMS to use the Z-score method and a 90th percentile benchmark to preserve the magnitude of relative performance, rather than to force the ranks as the percentile methodology does.

B. Recommendation 2: CMS should use the categorical results for the standardized ratio measures to account for the statistical uncertainty with numerical values.

KCP also continues to recommend that CMS use the categorical results for the standardized ratio measures that are publicly reported via Dialysis Facility Compare (DFC), rather than the numerical values. As noted on the legend to the standardized ratio measures on DFC, there is a significant amount of statistical uncertainty inherent in the measures due to the way they have been designed. DFC explains this uncertainty to consumers with the following legend.

⁶CMS, “Design for *Nursing Home Compare* Five-Star Quality Rating System” 9 (Feb. 2015).



The yellow area shows the “range of uncertainty around the facility’s ratio.” This uncertainty means that if a facility has a ratio of eight, the actual ratio could be between four and 12. DFC compensates for this uncertainty by categorizing facilities into one of three categories: better than expected, as expected, or worse than expected.

In ESRD Five Star, CMS currently uses the numerical value as the basis for scoring and ignores the uncertainty that the Agency has for years acknowledged and that continues to serve as the basis for public reporting on the individual measures. A better approach would be to use the categorical results from DFC, assign a numerical score based on those categories, and then add the points to assign a domain score. In other words, a facility whose performance is “better than expected” would receive two points; a facility whose performance is “as expected” would receive one point; and a facility whose performance is “worse than expected” would receive zero points.

We strongly urge CMS to adopt this recommendation, which recognizes the statistical uncertainty of the standardized ratio measures.

C. Recommendation 3: CMS should use star category boundaries that are based on actual variations in overall performance.

Third, KCP recommends that CMS define star assignment by using benchmark-driven boundaries that relate to actual performance, instead of relying upon preconceived and fixed percentages of facilities per star category. ESRD Five Star currently relies on an arbitrary determination of the number of facilities that must be assigned to each star category. Further, as we discuss in the following section, no accounting of inherent statistical measurement error has been applied. Under the CMS approach, 10 percent of facilities will always receive five stars and 70 percent of facilities will always receive three or fewer stars. Because the actual distribution of performance is asymmetrical, the distribution is not a normal bell

curve. Given this fact, the 10-20-40-20-10 percentage distribution of star assignments cannot be justified.

A more statistically valid and appropriate approach is to use boundaries defined by actual variations in overall performance. Specifically, we recommend that facilities with overall scores within one standard deviation above or below the average score receive three stars. Facilities with overall scores with more than one standard deviation above the average would receive four stars. Facilities with scores more than one standard deviation above the average in every measure domain would earn five stars. Facilities with overall scores more than one standard deviation below the average would receive two stars. Facilities with scores more than one standard deviation below the average in every measure domain would receive one star. The chart below shows the recommended cut points. The number of facilities in each star category would be determined by the performance on all of the measures across all domains.

	Base Criteria ⁷	“Plain English” Description
★★★★★	$d_1-A, d_2-A, \text{ and } d_3-A \text{ all } \geq C$	“The facility’s performance in <u>every quality domain</u> is better than expected.”
★★★★	$x - A \geq C$	“The facility’s <u>overall performance</u> is better than expected.”
★★★	$x - A \geq -C \text{ and } < -C$	“The facility’s overall performance is as expected.”
★★	$x - A \leq -C$	“The facility’s <u>overall performance</u> is worse than expected.”
★	$d_1-A, d_2-A, \text{ and } d_3-A \text{ all } \leq C$	“The facility’s performance in <u>every quality domain</u> is worse than expected.”
Not Rated		“The facility has insufficient data in one of the measure domains”

Using boundaries based on actual performance variation avoids the use of an arbitrary percentage distribution of stars. It also aligns the star ratings more directly to the underlying DFC data. The approach would provide more meaningful

⁷d1, d2, d3 = weighted average of Z-score within each domain
 X = overall score (weighted average across all 3 domains)
 A = average of X for all facilities
 C = cut point value for Star Ratings, expressed as 1 standard deviation from A

information to patients, their families, and other consumers than the current ESRD Five Star assignment method.

D. Recommendation 4: CMS should set the star category boundaries to provide statistical confidence in the final stars.

Finally, KCP recommends that CMS set the star category boundaries to provide statistical confidence in the final assignments. As we have described during previous meetings, there are several problems related to the data. While adjusting for statistical confidence for the final stars does not address all of these concerns, it is a step in the right direction.

Based upon our analysis, using a three standard error confidence interval would provide almost 100 percent certainty that a facility is assigned to the right category. Three standard errors is a common confidence interval used to provide high statistical certainty.

We encourage CMS and the ESRD Five Star TEP to adopt this recommendation to compensate for the statistical uncertainty around boundaries for the star cut points and to be certain that facilities are not mistakenly assigned to categories (which will allow CMS to better defend and explain the star assignments).

II. Impact of Recommendations

The chart below shows the percentage of facilities in the different star categories if CMS were to adopt the KCP recommendations (using January 2015 data).

	Criteria After Adjustment for Statistical Confidence	Number of Facilities	Percentage of Facilities
★★★★★	$d_1-A, d_2-A, \text{ and } d_3-A \text{ all } \geq -0.23$	1586	25%
★★★★	$x - A \geq -0.23$	2549	40%
★★★	$x - A \geq -1.18 \text{ and } < -0.23$	1330	21%
★★	$x - A \leq -1.18$	111	2%
★	$d_1-A, d_2-A, \text{ and } d_3-A \text{ all } \leq -1.18$	6	0.1%
Not Rated		725	11%

III. Conclusion

In sum, we strongly urge you to adopt the recommendations outlined in this letter, as well as the recommendations contained in our letter of February 13, 2015. As we have noted previously, KCP supports a transparent and easily understood approach to providing patients, their families, and other consumers with quality information. Such a system, however, must be valid and reliable. Our concern from the beginning about the current ESRD Five Star program is that it centers on a methodology that does not accurately reflect the actual performance of dialysis facilities and can mislead patients, their families, and other consumers. If adopted, the recommendations set forth in this letter – as well as our February 13th letter – would significantly address many of our concerns.

Thank you again for your willingness to engage with us on this important program. Kathy Lester will be in touch to schedule a meeting with your team to answer any questions about the recommendations and discuss their implementation.

Sincerely,



Edward R. Jones, M.D.
Chairman
Kidney Care Partners

cc: Patrick Conway, M.D., Acting Principal Deputy Administrator for Innovation & Quality, Director of the Center for Clinical Standards and Quality, CMS Chief Medical Officer
Kate Goodrich, M.D., Director of the Quality Measurement and Health Assessment Group
Elena Balovlenkov, R.N., Technical Lead for Dialysis Facility Compare
Joel Andress, Ph.D., Center for Quality Measurement in the Health Assessment Group