

**End Stage Renal Disease
Dialysis Facility Compare Star Ratings
Technical Expert Panel**
Summary Report

April 27 & 28, 2015

| | |
|---|----|
| End Stage Renal Disease Dialysis Facility Compare Star Ratings Technical Expert Panel <i>Summary Report</i> ... | 1 |
| End-stage Renal Disease Star Ratings Technical Expert Panel Summary | 4 |
| TEP Objectives | 4 |
| TEP In-Person Meeting | 4 |
| 1. Introduction | 10 |
| 2. Overview | 10 |
| 3. Preliminary Activities | 10 |
| 3.1 Environmental Scan and Literature Review..... | 10 |
| 3.2 TEP Charter | 11 |
| 3.3 Pre-TEP Teleconference Calls | 11 |
| 4. TEP Meeting..... | 11 |
| 4.1 Introductory Joint Session | 11 |
| 4.1a. UM-KECC Presentation, Yi Li, PhD..... | 11 |
| 4.1b. CMS Introduction of Risk Adjustment Policy, Joel Andress, PhD | 12 |
| 4.2 Public Reporting/Patient and Consumer Understanding Workgroup Day 1 Discussions | 12 |
| 4.2a. Opening Discussion..... | 12 |
| 4.2b UM-KECC Presentation, Joseph Messana, MD | 13 |
| 4.2c. Discussion of Standardized Measures..... | 13 |
| 4.2d Star Rating Distribution and Absolute vs. Relative Rankings..... | 14 |
| 4.2e. Potential Unintended Consequences and Need for Patient and Consumer Education | 15 |
| 4.2f. Discussion of Relevant DFC Measures..... | 16 |
| 4.3 Methodology Workgroup Day 1 | 17 |
| 4.3a Opening discussion | 17 |
| 4.3b UM-KECC Presentation, Christopher Harvey (Part 1)..... | 18 |
| 4.3c Discussion..... | 19 |
| 4.3d UM-KECC Presentation, Christopher Harvey (Part 2)..... | 21 |
| 4.3e Discussion | 22 |
| 4.4 Summary of the Joint Workgroup Discussion-Session 1 | 25 |
| 4.4a Different Considerations for Patients | 25 |
| 4.4b Measures to Consider for Retirement..... | 25 |
| 4.4c Patient Reported and Patient Experience Outcomes | 26 |
| 4.4d Methodological Considerations..... | 26 |
| 4.5 Public Reporting/Patient and Consumer Understanding Workgroup-Day 2 Discussions | 26 |
| 4.5a Day 1 Recap Discussions | 26 |
| 4.5b DFC Website Recommendations | 28 |
| 4.6 Methodology Workgroup-Day 2 Discussions | 30 |

| | |
|---|-----|
| 4.6a Summary of Day 1 Recommendations | 30 |
| 4.6b Discussion | 30 |
| 4.7 Summary of the Final Joint Workgroup Discussion Session (Day Two)..... | 34 |
| 4.7a Public Reporting/Patient and Consumer Understanding Workgroup Recommendations..... | 34 |
| 4.7b Methodology Workgroup Recommendations..... | 35 |
| 4.7c Final Joint Session Discussions..... | 36 |
| 4.8 Public Comment Period | 37 |
| 5. Potential Follow-up TEP Teleconference Calls | 39 |
| 6. Appendices | 40 |
| Appendix A: Technical Expert Panel Charter | 40 |
| Appendix B: Public Reporting/Patient and Consumer Understanding Pre-TEP Conference Call Minutes | 43 |
| Appendix C: Methodology Workgroup Pre-TEP Conference Call Minutes..... | 47 |
| Appendix D: Introductory Joint Session Presentation | 52 |
| Appendix E: Public Reporting/Patient and Consumer Understanding Workgroup Presentation | 92 |
| Appendix F: The Methodology Workgroup Presentation | 133 |
| Appendix G: Public Reporting/Patient and Consumer Understanding Workgroup Recommendations | 195 |
| Appendix H: Methodology Workgroup Recommendations | 200 |

End-stage Renal Disease Star Ratings Technical Expert Panel Summary

The Centers for Medicare & Medicaid Services (CMS) has contracted with The University of Michigan Kidney Epidemiology and Cost Center (UM-KECC) to convene a Technical Expert Panel (TEP) to review the Dialysis Facility Compare (DFC) star ratings methodology. The purpose of the project is to solicit input from patients and experts with relevant experience for review and evaluation of the Dialysis Facility Compare (DFC) Star Ratings methodology, and the online presentation of the DFC Star Ratings.

TEP Objectives

The objectives of the End-stage Renal Disease (ESRD) DFC Star Ratings TEP are described in the TEP charter provided (Appendix A) to TEP members prior to the in-person meeting. The TEP was tasked with using existing data and the experiences and expertise of its members to formulate recommendations for UM-KECC regarding the star rating statistical methodology, the measures used in the star ratings (to consider measures for retirement or future implementation), and the presentation of the star ratings on the DFC website.

TEP In-Person Meeting

The in-person ESRD TEP was convened in Baltimore, MD on April 27 and 28, 2015.

The TEP was organized into two workgroups: the Methodology Workgroup, and the Public Reporting/Patient and Consumer Understanding Workgroup. The workgroups jointly identified overall goals prior to breaking out and broke out into separate workgroup sessions. This allowed TEP members to then focus on issues specific to patient and consumer understanding of the star rating, or those related to the technical details of the current methodology.

The TEP consisted of individuals from the following areas of expertise or experience:

- consumer, patient, and family perspectives on treatment;
- biostatistics and statistics methodology;
- clinical treatment of ESRD;
- dialysis organization operation; and
- performance measurement and quality improvement.

Public Reporting / Patient and Consumer Understanding Workgroup Agenda Topics:

- Description of measures included in the star rating
- Discussion and Q&A on current methodology
- Use of symmetrical distribution, and appropriateness of current rating categories
- Appropriateness/utility of ranking based on relative vs. absolute performance
- Consumer demand for consistency in ranking across federal programs
- Review of potential unintended consequences of the star ratings
- Discussion of how the star ratings are conveyed, and how to best educate the public

- Interpretation of information on the DFC website

Methodology Workgroup Agenda Topics:

- Measure Scoring
- Measure Weighting
- Categorizing Facilities
- Comparisons and Alternatives

Topics for Both Workgroups:

- Consideration of possible quality measure topic areas for inclusion in the next iteration of star ratings for facilities;
- TEP recommendations for future iterations of the star ratings.

The following individuals participated in this TEP:

| Name and Credentials | Organizational Affiliation, City, State | Conflicts of Interest Declared |
|--|--|---------------------------------------|
| Paul T. Conway TEP Co-Chair, Public Reporting/Patient and Consumer Understanding Workgroup | <i>President</i> American Association of Kidney Patients (AAKP) <i>Board Member</i> Mid-Atlantic Renal Coalition (MARC) <i>Board Member</i> Polycystic Kidney Disease Foundation (PKDF) | None |
| Richard J. Cook, PhD, MS | <i>Tier I Canada Research Chair (renewal) of Statistics and Actuarial Science</i> University of Waterloo Waterloo, Ontario, Canada | None |
| Joanna Y. Galeas | <i>Patient/Patient Advocate</i> Renal Support Network | None |
| Joseph D. Karan | <i>Patient Advocate and Director of Education</i> National Kidney Foundation of Florida | None |
| J. Richard Landis, PhD, MS | <i>Professor of Biostatistics and Professor of Statistics</i> University of Pennsylvania Perelman School of Medicine, Philadelphia, PA <i>Professor of Biostatistics and Professor of Statistics</i> University of Pennsylvania, Wharton School, Philadelphia, PA | None |
| Franklin W. Maddux, MD, FACP | <i>Executive Vice President for Clinical & Scientific Affairs</i> Fresenius Medical Care Waltham, MA <i>Clinical Associate Professor of Medicine</i> University of North Carolina School of Medicine Chapel Hill, NC | None |

| Name and Credentials | Organizational Affiliation, City, State | Conflicts of Interest Declared |
|---|--|--|
| Allen R. Nissenson, MD, FACP | <p><i>Chief Medical Officer</i> DaVita HealthCare Partners El Segundo, California</p> <p><i>Emeritus Professor of Medicine</i> David Geffen School of Medicine at University of California, Los Angeles Los Angeles, CA</p> | Salaried employee and shareholder of DaVita Health Care Partners |
| Jane F. Pendergast, PhD TEP Co-Chair, Methodology Workgroup | <p><i>Professor of Biostatistics</i> Duke University Durham, NC</p> <p><i>Professor Emeritus of Biostatistics</i> University of Iowa Iowa City, IA</p> | None |
| John (Jack) Reynolds | <p><i>Vice President and Patient Ambassador</i> Dialysis Patient Citizens</p> <p><i>Member</i> National Kidney Foundation (NKF)</p> | None |
| Tonya L. Saffer, MPH | <p><i>Senior Health Policy Director</i> National Kidney Foundation (NKF)</p> <p><i>Commissioner</i> Montgomery County Commission on Health</p> | None |
| Chris Sarfaty, MSW, LICSW | <p><i>Patient-Centered Collaborative Care Coach</i> Coaching for Health Professionals</p> | None |
| Dylan S. Small, PhD | <p><i>Professor of Statistics</i> Wharton School , University of Pennsylvania, Philadelphia, PA</p> | None |
| Nicole Stankus, MD, MSc | <p><i>Associate Professor of Medicine</i> The University of Chicago Chicago, IL</p> | None |
| Catherine A. Sugar, PhD, MS TEP Co-Chair, Methodology Workgroup | <p><i>Director of Semel Institute Statistics Core</i> University of California, Los Angeles Los Angeles, CA</p> <p><i>Associate Professor, Departments of Biostatistics, Statistics & Psychiatry</i> University of California, Los Angeles Los Angeles, CA</p> | None |

| Name and Credentials | Organizational Affiliation, City, State | Conflicts of Interest Declared |
|--|---|--------------------------------|
| Sumi Sun, MPH | <i>Director of Applied Research and Data Analysis</i> Satellite Healthcare San Jose, CA | None |
| David M. White | <i>Chair on the Patient Advisory Committee and Medical Review Board Member</i> Mid-Atlantic Renal Coalition (MARC) | None |
| "Anonymous Patient A" (withheld by request) | Health care research worker at a health care company that works under contract with CMS (non-ESRD topic areas). | None |
| Contractor Staff | | |
| Yi Li, PhD | Director of Kidney Epidemiology and Cost Center and Professor of Biostatistics | None |
| Joseph Messana, MD | Collegiate Professor of Nephrology and Professor of Internal Medicine | None |
| Richard Hirth, PhD | Associate Director of Kidney Epidemiology and Cost Center and Professor of Health Management and Policy | None |
| Claudia Dahlerus, PhD, MA | Principal Scientist | None |
| Ji Zhu, PhD | Professor of Statistics | None |
| Christopher Harvey, MPH | Graduate Student Research Assistant, Biostatistics, School of Public Health | None |
| Zezhi (Zac) Zhang, MPH | Graduate Student Research Assistant, Biostatistics, School of Public Health | None |
| Karen Wisniewski, MPH | Research Area Specialist Senior | None |
| Natalie Scholz, MA | Research Area Specialist Senior | None |
| Cindy Liao, MS, MPH | Research Area Specialist Intermediate and Lecturer I in Psychology | None |
| Casey Parrotte, BA | Research Analyst | None |
| Tempie Shearon, MS | Research Area Specialist Lead | None |
| Jordan Affholter, BA | Research Analyst | None |
| Jennifer Sardone, BA | Research Analyst | None |

Additional Stated Conflicts of Interest

TEP members stated additional conflicts of interest not originally disclosed in the TEP nomination forms:

- Franklin W. Maddux is the Chief Medical Officer and Executive Vice President for Clinical & Scientific Affairs for Fresenius Medical Care; his salary is fully provided by Fresenius Medical Care.
- Allen Nissenson is the Chief Medical Officer for DaVita; his salary is fully provided by DaVita Healthcare Partners.
- Nicole Stankus is the Medical Director of a University of Chicago dialysis unit that was acquired by DaVita Healthcare Partners in 2010. Her salary is fully provided by the University of Chicago.
- Sumi Sun stated her salary is fully provided by Satellite Healthcare.
- Paul Conway stated that he is the president of the American Association of Kidney Patients (AAKP).
- “Anonymous Patient A” has no current conflict of interest. She did report what may be a perceived conflict of interest in the future if her company’s pending bid on an ESRD contract is awarded.

1. Introduction

This report summarizes the discussions and recommendations of the ESRD DFC Star Ratings TEP convened on April 27 and 28, 2015 in Baltimore, MD, as well as the pre-TEP teleconference calls conducted on April 13 and April 14, 2015.

2. Overview

The TEP was divided into two workgroups. The Public Reporting/Patient and Consumer Understanding Workgroup addressed a number of issues, including:

- the standardized measures;
- star rating distribution;
- absolute vs. relative rankings of facility scores;
- inconsistency in how health care providers are ranked across CMS public reporting programs;
- potential unintended consequences of the star ratings, and
- the presentation of star ratings on the DFC website.

The Methodology Workgroup was tasked with discussing the following topic areas:

- measure scoring;
- measure weighting;
- star rating categorization;
- missing measure values in facilities;
- impact of outliers;
- sensitivity analysis, and
- the framework for adding new measures.

Both workgroups discussed potential quality measure topic areas for inclusion or exclusion in the next iteration of star ratings for facilities.

3. Preliminary Activities

3.1 Environmental Scan and Literature Review

Prior to the in-person TEP meeting, UM-KECC presented the TEP members with published literature and other background materials related to the statistical topic areas of the star ratings methodology.

The topics covered by the literature included:

- Small facility adjustments
- Clustering methods
- Multivariate analysis
- Star rating calculations in other health settings.

Other materials provided to the TEP members included the Consolidated Questions Document that addressed community questions and concerns received and responded to by CMS, a technical user's guide on the DFC Star Rating, a guide to DFC reports, and a list of organizations that had commented on the DFC Star Ratings.

3.2 TEP Charter

The DFC Star Ratings TEP Charter was publicly posted with the Nomination Materials, and was distributed to the TEP members for review. The TEP Charter is included as Appendix A.

3.3 Pre-TEP Teleconference Calls

Two preliminary teleconference calls preceded the in-person TEP meeting. The pre-TEP conference call for the Public Reporting/Patient and Consumer Understanding Workgroup occurred on April 13, 2015. The pre-TEP conference call for the Methodology Workgroup occurred on April 14, 2015. The pre-TEP conference calls focused on the introduction of TEP members, the role of the TEP, the TEP Charter, and the process for accessing TEP resources for each workgroup. The two workgroups were provided with the pre-TEP teleconference minutes.

Both TEP pre-conference calls included a period for public comment. The pre-TEP teleconference minutes and public comments are included as Appendices B and C.

4. TEP Meeting

4.1 Introductory Joint Session

Following individual introductions and conflict of interest (COI) disclosures, a project overview was provided to all TEP participants. Yi Li, PhD, provided a summary of the current DFC Star Ratings. The presentation is available in Appendix D. Joel Andress, PhD, provided background on CMS risk adjustment policy. Dr. Andress is the CMS Contracting Officer's Representative for the ESRD Quality Measure and Public Reporting contract.

4.1a. UM-KECC Presentation, Yi Li, PhD

Dr. Li gave an introduction to the general star rating methodology to a joint session of both TEP workgroups. Key issues included the reported measures in the star ratings, measure scoring, measure domains, how to address missing measure values, and star rating cut-offs. Other background issues included the history of the DFC website, the federal mandates for increased transparency, and an overview of the Medicare.gov Compare websites and star ratings.

Dr. Li explained that star ratings are intended to assist patients in evaluating quality of dialysis care, and to ensure accurate reporting of facility outcomes for use by consumers. The star ratings are intended to supplement, not replace visiting a facility, speaking with a doctor, or looking up data on individual quality measures. Details of the presentation can be found in Appendix D.

4.1b. CMS Introduction of Risk Adjustment Policy, Joel Andress, PhD

Dr. Andress addressed both TEP workgroups to provide background information about CMS practices, specifically with regard to risk adjustment of measures for socio-economic status (SES). In the past, CMS has implemented other measures that were risk adjusted at the measure level; these were developed with the intention that the measures can be used for different program purposes. Dr. Andress also explained that they are currently participating in the National Quality Forum (NQF) two-year trial period for developing and testing quality measures that risk adjust for SES. CMS policy has generally been to not risk adjust, but that policy is currently undergoing review. The government is also in the process of producing a report on the impact of adjusting for SES at the measure and program level, across all federal health reporting programs. Dr. Andress noted that the report will not be completed in time for the next iteration of the star ratings.

Dr. Andress emphasized that CMS and UM-KECC cannot respond immediately to requests to adjust for SES as CMS is waiting on the completion of the report. Dr. Andress acknowledged that this is an area of concern, but recommended that it not be a primary focus for the present TEP discussions. He did note that TEP members would have opportunity to make general recommendations, including exploring risk adjustment for SES, in the time allocated. During the first day joint session, time was set aside for TEP members to raise any issues. He also assured participants that CMS intends to consider all input as policy decisions are made about future iterations of the star ratings.

4.2 Public Reporting/Patient and Consumer Understanding Workgroup Day 1 Discussions

4.2a. Opening Discussion

The Public Reporting/Patient and Consumer Understanding Workgroup began with additional introductions and an overview of the agenda and goals for the workgroup. Dr. Messina stated that the decision to separately discuss issues related to public reporting and consumer understanding and issues related to technical and statistical analyses was made to allow ample focus on the issues and opinions of patient and consumer advocates. This workgroup would also discuss the star rating methodology, but at a different level of statistical detail. The moderator then reviewed the agenda, noting that it was designed as an attempt to focus on some of the comments received from the renal community during the previous eight months. It was also noted that on the second day, time was reserved to discuss suggestions for best educating the public about the star ratings, and how information is currently presented on the Dialysis Facility Compare website. More than one member of this workgroup expressed the wish for the two workgroups to have more of an overlap, as the Public Reporting/Patient and Consumer Understanding Workgroup topics are inextricably connected to methodology.

To begin, Ms. Balovlenkov, DFC Lead for CMS Public Reporting, provided background information on the initial creation of the star ratings. Ms. Balovlenkov said CMS has received feedback from the community regarding issues of cognitive burden of the information on DFC and other Compare sites. Additionally, one of the goals behind the star ratings was to have one system across Compare sites that the public is familiar with and can identify with. She also stated that the goal of this TEP workgroup is to guide the future of the DFC measures and star ratings. Additionally, Ms. Pratt, Director DCPAC (CMS), noted that the existing data (individual measures) are still available on DFC; CMS just added another level of information (overall star rating) based on comments from the public.

4.2b UM-KECC Presentation, Joseph Messana, MD

Dr. Messana presented slides (see Appendix E) summarizing many of the topics arising from public comments, including:

- star rating distribution;
- absolute vs. relative rankings;
- consistency in ranking across public reporting programs;
- potential unintended consequences; and
- presentation of star ratings on the DFC website.

These topics were used as a basis for discussion among the members of the workgroup.

DFC Star Rating Standardized Measures

Dr. Messana gave a more detailed description of the DFC Star Rating standardized measures—the Standardized Mortality Ratio (SMR), Standardized Hospitalization Ratio (SHR), and Standardized Transfusion Ratio (STrR). The SMR, SHR, and STrR are all risk adjusted outcome measures, and were developed based on the input of previous clinical TEPs. It was explained that, for the standardized measures, a lower score is better. All standardized measures are risk adjusted for patient characteristics.

An overview was provided of each of the three standardized outcome measures, describing the outcome being reported, patients included in the measure, and patient characteristics included for risk adjustment.

4.2c. Discussion of Standardized Measures

Following the review of the standardized measures, the Public Reporting/Patient and Consumer Understanding Workgroup TEP members discussed the measure displays and specifications. One TEP member noted that they would prefer to see the information as a percentage, rather than as a standardized ratio, as it is easier to understand. There was also some concern from the community noted regarding the accuracy or completeness of the information that is provided on the CMS 2728 Medical Evidence forms; it may be important to examine the process of form completion for potential improvements. Some issues regarding the frequency of the completion of the 2728 form, and who it is completed for (e.g., patients returning to dialysis after failure of longstanding kidney transplants) were discussed. It was noted that CMS will be made aware of these concerns. It was also requested that the report note there were concerns about CROWNWeb data.

There was also discussion of accounting for socio-economic (SES) factors in the calculation of standardized measures. TEP members stated that discussing socioeconomic issues is important when attempting to explain the measures to patients. One TEP member noted that many dialysis facilities' outcomes reflect the SES of patients who dialyze there, so patients' SES is related to how the patients are cared for. One TEP member stated that it is important to clearly communicate the facility quality outcomes to patients, because many patients may feel that the measure values are a direct reflection of how well they individually fare in achieving the clinical outcomes assessed by the facility. Another TEP member stated that SES should be explored along with other demographic factors, but there may be a risk of explaining away the differences between patients by accounting for many different patient demographic and SES characteristics.

It was also noted that there is a need to support and inform the staff and facilities in their understanding the ratings and measurements because, although not everything can be addressed by the facility, facilities can improve how they educate patients on the measures. Ms. Balovlenkov (CMS) stated that the facility

staff training may vary based on the business model or company procedures of each facility. The variability in staff training may affect their approach to educating patients on the measures.

4.2d Star Rating Distribution and Absolute vs. Relative Rankings

To compare different types of distributions for reporting star ratings, TEP members were presented with graphs showing the current star rating categories (see Appendix E for slides). They were also shown examples of what the distribution of star ratings would look like if they were based on the calculation methodology used for determining the QIP payment reduction categories, or if they were based on equal quintiles (20% in each star category). It was also noted that the final score using the measures in the current star rating approximates a symmetric distribution.

Dr. Messina presented information explaining absolute vs. relative rankings for the star ratings, including examples of each scoring approach (see slides in Appendix E). TEP members felt that the specific methodology based on relative ranking was more complicated to understand. They provided feedback on which method they believed would be more easily understood by consumers and patients. TEP members also discussed whether it was preferable for all CMS quality reporting programs to use the same method for calculating overall quality, or different methods. For example, this was discussed in the context of the methods used to calculate overall scores for the QIP and the DFC Star Ratings.

TEP members stated that the current rating system has limitations due to its complexity, and that it is difficult to explain to patients. A TEP member noted that it is easier for a facility to explain to a patient the meaning of their lab values than explaining the star ratings to patients and how to interpret them. It was also noted that star ratings are based on the measures that are currently available on the DFC website, but these measures might not be as meaningful to patients as patient-reported measures.

One TEP member noted that for the current measures, it may make more sense to have absolute cutoffs for measure scores, but that a relative ranking system may be better once patient centered metrics are incorporated. There was agreement among the TEP members that they would prefer that measures be based on achievement of absolute thresholds/cut-offs rather than relative rankings. Some members of the TEP noted that absolute cutoffs are a good starting point, even though there is the issue of determining the threshold. They suggested improvements such as the ability to compare facilities to peer facilities (i.e., peer group comparison) to ensure patients are aware of what quality of care was provided in those similar types of facilities.

There was strong support for an interactive rating approach that allows patients and patient advocates to select measures/aspects of care that are important to them to be included in a star rating. An example several members used was a star rating for consumer services, such as for a hotel. For hotels, one can read the reviews of people who stayed at the hotel and find out information directly from the consumer. TEP members suggested it may be useful to do something similar with the DFC Star Rating, that is, include written reviews from patients that would accompany the star rating. TEP members also discussed having the ability to select the data on DFC that was most useful to the individual user. As a result of this discussion, the consensus of the patient group was that, where feasible, metrics based on absolute achievement were easier to understand and therefore, preferable for use on the DFC site (and in star ratings).

The Public Reporting/Patient and Consumer Understanding Workgroup expressed interest in adding one or more domains to the star rating that are based on patient experience of care. The TEP also discussed the appropriateness of the current star rating categories. Some members noted that it may be confusing to have different CMS program star ratings based on different measures. For example, Hospital and Nursing Home Compare have different measures specific to those care settings. There was not complete

consensus on this point. Some workgroup members recognized that consistency across different CMS Compare programs was probably not as important because they measure such different processes.

One TEP member asked CMS staff present at the workgroup to provide clarification on whether or not the DFC Star Ratings were intended to be a mechanism for determining facility reimbursement rates for providers. The CMS staff resource indicated that, at this point, that was not the intention of the DFC Star Ratings. One TEP member stated that it is unrealistic to think that all of the different measures can be truly consistent across programs because the measures in different clinical areas would be different than the clinical measures used for reporting quality in dialysis facilities. It was clarified that the desire wouldn't necessarily be to make all of the ratings consistent in terms of the same measures, but that it is confusing to have data based on absolute achievement (thresholds) reported on one site and percentages reported on another site and then star ratings using different calculation methodologies for the different Compare sites' star ratings.

It was also noted that the star ratings could be useful if defined based on what patients and consumer advocates find important. It was also noted that since the rating is provided on a government site, it carries more weight and credibility than non-government ratings sources. The TEP workgroup discussed the lack of QIP knowledge among most patients. It was stated by a TEP member that the QIP itself is rarely discussed with the patient community. It was stated that facilities are required to educate patients on the QIP so that's why there may be some basic knowledge for some patients about the QIP. Because of this some workgroup members felt it might be more important to have consistency between measures and overall scores reported in DFC and QIP. The TEP workgroup discussed the lack of QIP knowledge among most patients. It was stated by a TEP member that the QIP itself is rarely discussed with the patient community.

4.2e. Potential Unintended Consequences and Need for Patient and Consumer Education

The group next discussed the potential unintended consequences of the star rating. Many of these had previously been identified through comments submitted to CMS regarding the current star rating program. The TEP discussed potential unintended consequences, such as:

- involuntary discharge of patients;
- "cherry picking" patients;
- potentially creating false differences between facilities;
- the contribution and impact of geographic variation/disparities;
- patient concerns with the ratings of their local facility choices; and
- the lack of perceived quality alternatives within easy commuting distances.

TEP members identified a possible additional consequence: low star ratings could have a negative impact on the morale of both the patients and staff of a facility. It was stated that it may reduce the power that individual patients feel over their own care, including the possibility of patients trying to assure themselves that "it's not my fault; it's the facility's fault". TEP members felt that education is needed when implementing the star ratings in a way that reduces this possibility.

TEP members provided feedback on the best methods for educating the public about the DFC Star Ratings, including that the star ratings should be explained in plain language. They also gave input on the level of content detail that may be useful in outreach and education methods. TEP members reviewed and gave feedback on the display of the star ratings on the DFC website, and how they believed the star ratings should be used by consumers.

TEP members stated that patients and facility staff should be made aware of how they can contribute to improving their facility's rating. Some TEP members noted that it is an important goal to encourage patients' buy-in and engagement, to be a driver of change in facility care. They were unsure, however, of how realistic it is to put that responsibility on the patient. For example, patients tend to only go to the DFC site when they are curious or need immediate information.

The TEP stated clearly that the need for further patient education is very important. It appears that most patients select their facilities based on their physician's recommendation. A TEP member also noted that the intent of developing the star ratings was to make the process of selecting a facility easier for patients. One TEP member recommended using a short, directed video to display at the clinic. The video would educate patients on the star ratings while they are dialyzing. Ms. Balovlenkov stated the ESRD networks could potentially help produce an educational video.

One TEP member stated that the star ratings were intended to be transparent, easily understood, and useful to patients. TEP members believed that the language explaining the current methodology is too difficult to understand. Another TEP member noted that clinics will always need to continue educating patients about the clinical measures, what they mean, and how they are actionable.

4.2f. Discussion of Relevant DFC Measures

The TEP was asked to identify which current measures and information presented on the DFC website are meaningful to patients and patient advocates, and for their recommendations for new measures. The workgroup suggested patient satisfaction surveys, referencing how information from patient peers is important. They also said they would like patients to be able to develop a personalized rating of facilities by selecting their own criteria/data, providing them control and a better understanding of the information used to determine the rating. A TEP member noted that important information included which modalities were offered at a facility, shift times, and information on patient safety measures, such as patient injuries, falls, and cleanliness. They also mentioned grievances, including grievances filed to internal committees, CMS, ESRD Networks, or their state. Quality of life measures were mentioned, as well as treatment timeliness of the facility (time from being put on the dialysis machine to the time they are taken off) and attention to detail were also mentioned. Information about travel experiences and feedback from incoming transient patients were noted as items that would allow them to better review others' experiences at a particular facility. Another item mentioned the importance of knowing whether patients experience cramping, and how the facility handles it; one TEP member noted that this is essential to the patient's experience and impacts how the patient feels. Another item mentioned was reporting of "adequate" staffing. One TEP member noted that adequate staffing is important because if there is not enough help or if the staff is working too many hours, mistakes can occur. Sometimes mistakes happen, but reports of incidents are often not filed by patients for several reasons. Another topic discussed was facility and staff receptivity to issues that patients bring to their attention. It was noted that when staff convey they are listening it shows respect for the patient, and having this type of information about a facility would be useful. Happiness and quality of life measures were mentioned again in the discussion. Also, modality-specific measures that PD patients and home hemodialysis patients would consider as important were suggested by TEP members. Other suggestions regarding home dialysis therapies included information on time spent with the patient, education and time spent on educating the patient, information on patients following their care plans, infections of PD patients, social worker and nutritionist staffing levels, and supply reliability (dialysis supplies) or after hours support.

One TEP member presented results from a survey conducted by their organization, asking patients how they judge the quality of a facility. The highest number of respondents indicated that they judged quality based on the attention they received by the dialysis facility staff. The other responses selected by patients, in order of greatest choice, were 1) the look and comfort of the facility, 2) lab results, 3) how

they feel daily after dialysis, 4) how often they see the doctor, and 5) how long they have to wait to get an appointment. Respondents also judged quality based on how they viewed the skill set of the facility staff.

More survey results from the National Kidney Foundation showed that 69% of patients surveyed receive dialysis treatment in-center, and that 87% do not use online tools. Therefore, there is a need to promote greater use of online tools. One TEP member suggested several strategies; knowing the demographics of the majority of people that use social media could inform these strategies to engage more people in using online tools. These suggestions were further discussed by the TEP. Ms. Balovlenkov stated that CMS is aware that many patients do not have personal computers, but the system was created to allow access by a cell phone with internet.

The TEP then discussed what measures could be potentially removed from DFC. The TEP members suggested transfusion and hypercalcemia. One TEP member reported that they were not interested in mortality, hospitalization, or transfusions unless the event was directly caused by the facility. Other TEP members stated that they are very interested in mortality, and it should remain on the site.

4.3 Methodology Workgroup Day 1

4.3a Opening discussion

The Methodology Workgroup began with a brief discussion by TEP members of the mechanisms for adding new quality measures to DFC. CMS explained that to date there has been no formal systematic process for adding measures to DFC. CMS also noted that all measures being considered for the star ratings require a period of public reporting before being included. Therefore, the only measures that can be added in the short term are those that are already being reported, but are not being used in the star ratings calculations. Dr. Andress described the overall three-year review process using the example of SMR and additional potential comorbidity data sources that will be evaluated by an upcoming TEP with the goal of considering these data for an augmented risk adjustment model for the measure. There was also discussion and a request for clarity about how measures get added to DFC. In terms of recommending new measures, it was identified that one of the foci of the Public Reporting/Patient and Consumer Understanding Workgroup is making recommendations on new measure topics that are important to patients. Given the desire to add new measures in the future, one additional consideration for the star rating methodology is that it's capable of scaling to new topics.

The TEP members listed a number of issues that they hoped to cover during the TEP meeting, including:

- Concern that reporting the standardized measures as “better, worse, or as expected” is confusing to patients and providers.
- Concern that the measures grouped into each of the domains do not make intuitive or clinical sense, thereby leading to confusion.
- The ranking that occurs in the current star rating methodology is confusing to patients, providers, and observers of the renal community.
- Measure specificity is not sufficiently granular to allow for valid inter-organizational comparisons. The business rules by which people operate to collect data on these metrics need to be exactly the same. There needs to be clarity in data methodology, particularly for biomedical metrics.
- Lack of consideration for geographical variation.
- Create a disclaimer on the website that the data being used is not contemporaneous (based on data from 2013) and current performance may differ.

- Descriptors listed next to the stars are problematic (“below average” facility). They may be technically accurate, but may negatively impact patient perception/experience.
- Scoring based on performance and facility improvement would help drive improvement, and would align more closely with the QIP. There is some confusion in the community regarding the calculations used in the various public reporting programs.
- Consideration of the impact of the star rating methodology on small facilities.

UM-KECC noted that these were all important issues, and that several of them were discussed during the UM-KECC presentation. All of these topics were discussed during the Methodology Workgroup discussions, although consensus was not reached on several of the topic areas. The Methodology Workgroup recommendations and main topics of discussion are reflected in section 4.7b.

4.3b UM-KECC Presentation, Christopher Harvey (Part 1)

The Methodology Workgroup began their breakout session with a presentation by Christopher Harvey, Graduate Student Research Assistant, of UM-KECC. The presentation addressed many of the discussion points, including measure scoring, measure weighting, star rating categorization, missing measure values in facilities, outliers, sensitivity analysis, and the framework for adding new measures. Following Mr. Harvey’s presentation (see Appendix F), the Methodology Workgroup discussed the different topic points and approaches.

The presentation was organized by the three decisions that systematically constitute the framework of star rating—measure scoring, measure weighting, and star rating categorization.

Decision 1: Measure Scoring

Mr. Harvey gave an overview of six different methods that can be used for measure scoring, including:

- Minimal Transformation: measures are only adjusted in direction (a higher score is better) and scale (i.e., all measures range from 0 to 100);
- Percentile Ranking: ranking quality measures on a uniform distribution between 0 and 100 (i.e., the same number of facilities are given each value);
- Probit Ranking: ranking measures on a normal distribution between 0 and 100 (i.e., more facilities are given a middle value, as opposed to the percentile ranking);
- Clustering: identifying groups that contain values that are more similar to each other, and less similar to values in other groups;
- Percentile Thresholds: grouping measures based on their relationship to national averages;
- Performance Thresholds : grouping quality measures based on the fixed values of the measure; and
- Z-Score: a centering method based on how many standard deviations a measure’s value is away from the mean of that measure.

In general, as the data show, ranking methods tend to control the impact of outliers as the rankings do not depend on the distribution of the original measures. Ranking also ensures that measures exert the same influence on the rating despite the original distribution and scale of the measures. Part of the TEP discussion could include weighing the advantages of probit ranking (controlling outliers, giving measures equal influence) and z-scores (preserving measure distribution, but may be sensitive to outliers).

Decision 2: Measure Weighting

The second decision involves considering the measure weighting options, such as equal weighting, importance weighting, and adjusting for redundancy. Equal weighting assumes all measures should have the same influence in the rating, and therefore should be weighted equally. Importance weighting would follow a method based on expert opinion of relative value of the measures. However, there is currently no established consensus for the importance of measures in dialysis patient outcomes. Finally, adjusting for redundancy attempts to avoid double counting aspects of quality. With the aid of factor analysis, groups of measures are formed based on correlations, and groups are equally weighted.

Currently, STrR, SHR, and SMR are grouped into one domain. The Kt/V and hypercalcemia measures are grouped into a second domain, and the fistula and catheter measures form a third domain. The probit ranked measures are equally weighted within a domain, and the domains are equally weighted to create a final score.

Decision 3: Star Categorization

The third decision is considering the various star categorization options. These options include percentile thresholds, quality thresholds, final score clustering, and average quality measure star ratings.

The percentile thresholds fix the annual proportion of facilities in each star rating category. The quality thresholds fix final facility scores in each rating, or require certain scores on each measure or group of measures to attain the rating. A final score clustering approach groups final scores with statistical clustering, so that groups contain values that are more similar to each other, and less similar to values in other groups. The average quality measure star ratings approach numerically rounds the star ratings created for individual measures.

For percentile thresholds, fixed deciles (e.g., 10% 1-star and 5-star, 20% 2-star and 4-star, 40% 3-star) were chosen for the star rating. Fixing top and bottom performers may be problematic if the distribution of facility scores changes over time. For quality thresholds, fixing final scores may be difficult because standardized measures are relative to the scores of other facilities in that year. Quality thresholds would fix measure value cut-offs, which essentially groups individual measures and results in loss of information. For final score clustering, different clustering methods can produce different results. Outliers can form their own clusters. For average quality measure star ratings, fixing the measure value of grouping cut-offs results in a loss of information.

4.3c Discussion

One of the TEP members asked CMS to comment on the use of “average” in the definition of the star ratings, as compared to tying the ratings to clinical outcomes (like an absolute threshold). CMS explained that the concept of a threshold sounds appealing, but setting an appropriate threshold for each particular measure would require a process to arrive at a consensus, would require regular updating, and would be problematic to implement. CMS stated that they might investigate alternatives to the concept of comparisons to the average, but the current method did seem like the best option for the initial version of the star ratings.

Several TEP members went on to state that they would be open to the idea of continuing to use the average, should there be a more clear interpretation of what it means. The TEP also acknowledged that the measures currently displayed on DFC may not be of the most interest to patients—as one TEP member explained, what appears to matter most to them are patient-centered issues such as continuity of care and safety. CMS acknowledged this concern, and noted that quality measure development in ESRD is fairly new. In addition to those quality measures currently reported on DFC, there are additional

measures that are under consideration for inclusion in the star ratings that will provide more of this type of information to patients (such as the Standardized Readmission Ratio, In-Center Hemodialysis Consumer Assessment of Healthcare Providers and Systems, etc.). One of the primary purposes of the Public Reporting/Patient and Consumer Understanding Workgroup is to solicit feedback from DFC consumers on the addition of potential new measures.

Measure Scoring and Small Facilities

The TEP chairs asked the other TEP members to provide preliminary comments on the three decision points covered in the presentation. One TEP member asked if uncertainty in the standardized measures is accounted for when calculating the standardized measures, as this may be an approach to help stabilize the results for small facilities. UM-KECC explained that they did not implement this in the methodology, and only reported the point estimates on DFC. UM-KECC explained that fully accounting for uncertainty in the star rating requires the use of patient-level data and resampling methods, such as the Bootstrapping method.

One TEP member asked UM-KECC to explain what they see as the pros and cons of the ranking method. UM-KECC explained that ranking seems easier to understand, and is more robust towards the outliers. Rankings are also more comparable across measures, since one can operate rankings on the same scale. UM-KECC was also aware of the cons to the rankings, including the fact that it ignores the physical distance between the real values. When using percentile ranking, the original distribution is effectively changed to a uniform distribution. To compensate for this, UM-KECC applied the probit transformation, which places most facilities in the middle unless they are truly extreme.

One TEP member noted that distance can be very meaningful, and the facilities that are in the tails are the most important. UM-KECC explained that they wanted the methodology to reduce the impact of outliers (as distributed in the tails) which are typically produced by small facilities, while identifying facilities with issues. For example, if a facility is performing poorly, with every measure value on the (bad) tail, then the probit transformation will detect this while using the original values which fall below average. Another TEP member added that ranking preserves the relative information, as it does not depend on the original values. A con of ranking is that it may lose the actual score, and may not be directly helpful for clinically significant interpretations.

UM-KECC was asked if there were any data available on the reliability/stability of the different measures. They explained that stability depends somewhat on facility size. If a facility has a very large number of patients, then its estimate will be very precise and will be stable across years. One TEP member noted that this gets back to the uncertainty rating—if the facility is very small, to claim that the methodology can accurately give a rating may be unrealistic. The group wondered about the number of small facilities, and how much weight they should be given when making scoring decisions.

CMS explained that there are many small facilities. The average number of patients in a dialysis facility is 76 patients, which is relatively small for many of these measures. A minimum number of patients or events are required to report each measure (11 patients for the non-standardized measures and a corresponding number of events or patient years at risk for the standardized measures). The TEP expressed an interest in dividing facilities by size, applying the methodology to each group (small, medium, large) to see the distribution of the star ratings. UM-KECC presented findings on the distribution of star ratings by facility size during the next methodology presentation.

4.3d UM-KECC Presentation, Christopher Harvey (Part 2)

Methodological Issues

Mr. Harvey provided an overview of additional methodological issues for the TEP to consider, including missing measure values in facilities, facility size adjustment, and the framework for adding new measures (see Appendix F). He also presented an overview of recommendations that had been provided by the dialysis community.

Comparison of Methods

Mr. Harvey presented a graph showing the differences in final scores using the probit ranked measures versus z-scored measures. The z-score method has more extreme values on the lower end of the range, and is more sensitive to extreme measure scores. Extreme scores may reduce or elevate a facility's ranking. Facility ratings are often diminished by scoring poorly on the skewed measures, of Kt/V and Hypercalcemia. The probit ranked measures and z-score ranked measures have similar distributions in relation to the standardized measures and intermediate measures. Only seventeen facilities differed by two stars when comparing the two different methods.

Stability Analysis

Probit and z-score produced nearly identical results: 54% of facilities are given the same rating from year one to year two using probit ranking, while 55% of facilities receive the same rating from year one to year two using the z-score method. This is compared to the probability of 26% for facilities to receive the same year-to-year rating by random chance alone.

Recommendations from the Community

UM-KECC presented a summary of the suggestions that have been submitted by the dialysis community. These included creating domain/measure thresholds necessary to obtain a 1-star or 5-star rating, scoring measures based on confidence intervals, and shifting star-rating cutoffs based on confidence intervals. The z-score and probit method were discussed, but no workgroup consensus emerged on this issue. Other methods were not discussed in detail.

Missing Measure Values in Facilities

UM-KECC explained that 9% of facilities are not rated, and the rate of missing values is about 10-20% for each measure. Currently, facilities missing a domain are not rated. If no domains are missing, missing measures are imputed with the national average rank (probit rank of 50). The current method shrinks towards the average. TEP members discussed the appropriateness of this method, as noted in section 4.3e.

Facility Size Adjustment

One consideration supporting the use of probit transformation is to compensate for excessive variation of the measures from small facilities. Small facilities tend to have more variation in their measures and thus may be overrepresented in the extreme star rating categories. UM-KECC presented a graph showing the distribution of star ratings by facility size. The probit ranking (current method) gives more balanced 5-star ratings than the z-score method, which gives smaller facilities disproportionately more 1-star ratings than the probit ranking method.

Framework for Adding New Measures

The star rating methodology requires flexibility in order to accommodate the addition or removal of measures. The measure domains can therefore be reevaluated or redesigned with each new iteration, including as new measures are added. One future consideration is to avoid having a single measure in a domain. It may give that measure too much influence on the rating, or, if that measure is missing, it could result in many facilities not receiving a rating.

4.3e Discussion

Measure Scoring and Small Facilities (continued)

A question for the TEP to consider is the rating of small facilities. Several TEP members raised the importance of considering the other common characteristics of small facilities, such as new facilities, and those in rural areas. These characteristics could be confounders.

The TEP generally agreed with including more years of data in the star rating calculation to limit the impact of annual variability. Currently, only the SMR uses data from more than one year. This may also present an opportunity to incorporate improvement, though it is still unclear about the best way to implement it. For an improving facility, one TEP member explained that an increase in its facility score may depend on whether recent or historical data is more heavily weighted.

Another option to stabilize the ratings for small facilities is to “shrink” to the average. One TEP member felt that shrinking the missing measures to the national average can be misleading, as the facility might actually be performing much better or much worse than average. This issue is difficult to convey within the transparent context of public reporting.

One TEP member asked if facility size was the greatest influence—either statistically or clinically—for a facility’s rating. Providers were asked to weigh in on other possible stratifications that could be examined. It was explained that there are dozens of characteristics that can be used as stratification options, such as rural/urban setting, teaching/non-teaching facility, and nursing home settings. The group then discussed the extent to which all facilities should be aiming for the same outcomes, regardless of their case mix. One TEP member noted that they would not be in favor of making star ratings based on statistical clustering, as that would be very confusing to patients.

The discussion next focused on geographical issues, particularly the concept of comparing facilities by region and/or incorporating a geographic adjustment. For example, 95% of patients dialyze within five miles of their home; in a rural area, there are few such choices for dialysis facilities. CMS explained that whether or not to account for geography is a policy decision. This issue has arisen for some other quality measures in the past, and it was determined that CMS preferred to not adjust for regional differences in the practices of care.

One TEP member pointed out that if the goal of the program is for patients and families to make decisions locally, then it would make sense to show the comparisons based on region. CMS explained that if the dialysis facilities in a particular region are not performing well compared to other facilities in the nation, which is valuable information, especially for patients who may travel. The current DFC website does support comparing facilities by region (searching based on ZIP code). CMS also noted that other members of the public view the website, and do view the data nationally.

Weighting, Grouping, and Framework for New Measures

The TEP chairs explained that the current methodology weights all measures equally. There are statistical considerations for discussion, such as giving measures less weight if they appear to be unstable or not measured accurately. Weighting based on clinical value may be difficult. The TEP chairs recognized these issues, but noted that the TEP is not likely to come to a consensus about the weighting method in the current session.

One TEP member asked UM-KECC if they knew how many facilities would receive five stars or one star on each of the seven measures. The TEP member stated this could indicate which domains are more influential. UM-KECC replied that if star ratings were given for each measure, assuming the current distribution cut-offs, only five facilities would earn five stars on all measures.

One TEP member wondered if the grouping of measures into the three domains has significant impact. Another TEP member explained that the current methodology uses factor analysis to identify which measures are the most similar. Factor analysis indicated that the best approach to the set of seven measures was to divide them into three domains, though the correlations within each domain were not high. However, the current method does not rely on “factor loadings”, and is essentially an equal weighting. As a result, the ratings actually could have been calculated without the creation of domains. The method does provide a framework for accommodating the new measures added in the future that would fall into different domains. That said, this TEP member stated that it would be helpful to have the domain concept or groupings already in place to ensure stability of the results.

The group went on to discuss the importance of considering clinically relevant groupings. One member urged the TEP to consider grouping clinically to help with patient understanding. While acknowledging that concern, another TEP member stated that it is important to determine if there is a statistical distinction between domains. If the measures are grouped into domains from a clinical perspective, but the measures within domains are not strongly correlated and the domains are not statistically separate, it can create analytical problems. The process for designing domains would likely need to respect both statistical and clinical needs.

Measure Scoring: Probit Ranking vs. Z-scores

The group then discussed the comparison between using a probit ranking and a z-score. One TEP member noted that they did not like the fact that probit ranking fundamentally changes the shape of the distribution (compared to the z-score, which maintains the shape exactly), and further stated that forcing “topped-out” measures (where a high number of facilities meet the measure) into a probit ranking creates a sense of futility for providers. Another TEP member replied that, in general, there are natural tensions between ensuring the combination of measures make sense, and respecting the original shape of the distribution.

One TEP member proposed a simulation that could be performed to weigh these different options. For this simulation, one would generate centers of varying sizes. The simulation scheme includes: pre-specify everything, generate the data, create the scores, do a transformation, rank them, and see how well it agrees with the “true” ranking. This would allow for an objective review of the effects of using each of these ranking systems.

Another TEP member pointed out that since this is a problem that is inherently multi-dimensional, it is difficult to know what the “right” ordering means. A rule should be stipulated to decide if, for example, facility A is better than facility B, otherwise we are comparing different and conflicting rules. In response, a TEP member pointed out that any one of the DFC measures inherits a natural ordering. For a facility to be ranked higher overall, it needs to have a small sum of ranks from these measures. Unless we can show

that one of the measures is more important, the overall ranking seems to be the best way to order the facilities. Another TEP member agreed, and added that the current methodology is reasonable and very data driven on this point. The statistical advantage is that it is relatively insensitive to which scaling metric (whether using probit transformation or z-scores) is used, and therefore can combine measures with different scales to reach an overall score.

Missing Data/Facility Size Adjustment

The group discussed the current method for handling missing values. If a whole domain is missing for a facility, then they are not assigned a star rating. If only one measure in a domain is missing, the value is imputed to the national average.

The TEP generally agreed that the current imputation method could be strengthened. One TEP member stated that there is probably not a uniform rate of missing data across all facilities (which can create some noise) and imputation to the national average may not be the best choice. The tradeoff of various imputation methods lies in complexity versus precision versus ease in understanding for the consumer.

As an alternative, a model-based imputation could be considered. A TEP member proposed regressing the missing value on the other facility variables—in other words, predicting the measure score for this facility based on their other values. Another TEP member explained that the general rule is that an imputation is more accurate with inclusion of as much information as possible. The group will need to decide which factors would be most influential.

Another option suggested by a TEP member and echoed by several others was a method referred to as the “nearest neighbor” approach: if a measure value is missing (e.g., hypercalcemia) in one facility, we can look at a certain number of other facilities with similar characteristics and possibly with similar other measure scores. These facilities are the so-called nearest neighbors. The average of these neighbors’ values on, for example, hypercalcemia will then be used to impute for the facility with missing measure of hypercalcemia.

Members of the TEP expressed interest in trying to demonstrate whether the imputations are accurate for each of the proposed methods. UM-KECC explained that they can simulate this by using the technique of cross-validation. They proposed taking a set of facilities’ measure values, artificially removing some results, and then testing each imputation option, to determine if the results match the measure values that were removed. The TEP agreed to this proposal, and was very interested in seeing the results of this simulation.

Small Facilities

To help interpret the star ratings of small facilities, one TEP member proposed providing supplemental information on the DFC website regarding the individual measure scores that contributed to the star rating, including indicators of uncertainty. The group also revisited the issue of providing historical data to help make the measures more reliable for small facilities. CMS understood that there is a conflicting concern—sufficient data are needed for valid measure calculation, but the usefulness of the data is limited by how current it is. Data currently used for the ratings calculation are already lagged because of Medicare claims delays and CROWNWeb reporting requirements and timing.

One TEP member asked if it was possible to determine (statistically) if there is a minimum number of patients required to make a measure calculation reliable. CMS noted that for the QIP, there is a small facility adjustor for facilities with 11-26 patients (for the standardized measures, this is expressed in number of events) based on the notion of inter-unit reliability.

Another TEP member explained that can be difficult to collect hard data to answer these questions, but simulations can be performed to show how a measure changes as the facility size grows. Unless one includes facility size as a covariate in the shrinkage model, small facility ratings will shrink back towards the center. This may be misleading to patients if notably good or notably bad facilities are pulled towards the middle.

The group requested that UM-KECC provide more information regarding the size of the facilities that score in the tails of each measure, to determine if there is a pattern. UM-KECC agreed to investigate this.

Disparities

The group briefly touched on patient disparities. It was noted that Large Dialysis Organizations (LDOs) are disproportionately located in rural and underserved locations. One TEP member asked if one would expect geographic differences in practice patterns. Another TEP member explained that there are some issues that are handled differently—some are provider driven, and some are population driven (by SES, cultural differences, etc.).

Cluster Analyses

One TEP member gave a brief summary of the clustering method, which tends to work best when there are areas of real concentration, i.e., concentrated areas of data with empty space in between. In that case, the distribution would show multiple “hot spots” (peaks) and is indeed a multimodal distribution. However, with only a unimodal distribution (with a single peak), even though some of the measures are skewed, there is actually no strong evidence of bunching. One can still use clustering to create a partition, but these techniques (e.g., K-means and Hierarchical clustering) do not appear to provide meaningful results.

4.4 Summary of the Joint Workgroup Discussion-Session 1

At the end of the first day of the TEP, both workgroups convened for a joint session. This discussion covered a number of topics, including measures that are appropriate and useful for inclusion in the star ratings.

4.4a Different Considerations for Patients

One TEP member explained that different patients have different goals. For the patient, the most important considerations are their own health goals and the care they receive. Some patients have the goal of receiving a kidney transplant, in which case they can consider the transfusion ratio and other clinical measures to give them information about facility care. Some patients prioritize a good quality of life and want to be able to walk out of their dialysis sessions and be active. Other patients may not want a transplant, but want to live as long as possible. Patients would choose a facility differently based on the particular performance in whatever outcome they are most interested in. A “cafeteria” approach may allow patients to focus on the measures that are most relevant to them.

4.4b Measures to Consider for Retirement

The group progressed to a discussion of the hypercalcemia and Standardized Transfusion Ratio measures. These were identified by the Public Reporting/Patient and Consumer Understanding Workgroup TEP members as possible measures to remove from the DFC Star Ratings. There was discussion of topped-out measures (e.g., hypercalcemia; dialysis adequacy as measured by Kt/V). One member of the Methodology Workgroup recommended caution on retiring metrics that are topped out (i.e., distribution is very close

with almost all facilities having very good performance), because there is a concern of facilities losing focus on maintaining excellence in basic quality care. The rationale for removal of metrics may be different for the hypercalcemia measure, because there is no clinical evidence to support the threshold of 10.2. Another TEP member stated that transfusions are major events that are important to understand, as there is concern that transfusions may eliminate patients who could be eligible for transplant. The TEP member asked if facilities and nephrologists should be held responsible for transfusions. UM-KECC stated that the rationale for the measure is that there is shared responsibility for outcomes in relation to transfusion. One TEP member stated that a large number of transfusions happen in the hospital and not in clinics. Clinics may not have a way to obtain that information accurately.

4.4c Patient Reported and Patient Experience Outcomes

There was general interest from several TEP members in finding a way to collect data regarding patient experience and patient-reported quality of care that could be incorporated in the star ratings. One TEP member's organization surveyed patients to determine what information is important to them. For example, quality of life was considered more important than the length of survival. Other important issues that were identified in the survey results include: how patients are treated, do the dialysis staff care about the patients, and if the facility staff treated patients with compassion. This TEP member's organization did a study several years ago that showed two main factors were correlated with patient survival: whether the facility staff did a medication review upon return from the hospital and the quality of education provided to the staff in dialysis facility. The challenge is creating a metric that measures quality of education in a dialysis facility. While the interaction of staff with patients is important, calculating only the number of staff is not necessarily related to quality of interactions for a facility.

4.4d Methodological Considerations

Additional discussion included a suggestion from members of the Public Reporting/Patient and Consumer Understanding Workgroup, explaining their interest in adding one or more domains to the star rating that are based on patient experience of care. Several TEP members acknowledged there will be additional methodological issues to consider if such measures are added.

One TEP member stated that the Public Reporting/Patient and Consumer Understanding Workgroup spent a lot of time discussing the methodology. There was a preference for having star ratings based on absolute performance (i.e., achievement of a set threshold such as the QIP) rather than relative ranking (i.e., relative performance such as the current star rating methodology).

A member of the Methodology Workgroup indicated that the methodology employed appropriately addresses the original CMS request given to methodologists for quality star ratings. Several TEP members expressed interest in adding regional benchmarking to the information provided with the star ratings, given that patients may be deciding between seven to ten local facilities. There was also interest in incorporating improvement into the star ratings to show whether the facility is performing better than it did in previous ratings.

4.5 Public Reporting/Patient and Consumer Understanding Workgroup-Day 2 Discussions

4.5a Day 1 Recap Discussions

The session began with restating that a TEP summary report will be written subsequent to the in-person meeting, and that it will reflect the TEP's comments and discussion during the two-days. A draft of the

report will be distributed to TEP members to ensure that it is an accurate summary of their feedback, comments, and discussions that took place during the TEP.

When discussing the re-cap of the first day, a Public Reporting/Patient and Consumer Understanding Workgroup TEP member noted that in general, the overall perceived relevance of DFC measures for patients/ consumers is low. Another noted that establishing standards (i.e., absolute standards or thresholds) was important to the group. Other members of the workgroup agreed that having performance standards was important. One TEP member expressed a preference for gradations of standards to assess performance rather than a pass or fail for the measures. This would allow the facilities to earn credit for partial success in working to achieve an absolute standard, allow facilities room for improvement, or at least let them explain why they have not improved. One TEP member offered the suggestion of allowing facilities to comment on their measure and star rating scores. The facilities would potentially allow consumers to access information that may explain why that facility received the certain star rating. The idea of adjustment for factors (such as socioeconomic factors) outside of facilities' control was again mentioned by members of the workgroup.

The group also discussed the complexity of the current methodology, stating that it is difficult to understand. It was noted that an issue is that this rating system is different from what some users are used to seeing in other consumer rating programs. It was stated that what really needs to be accomplished is a way to allow for patients and consumers to directly provide feedback. The idea of changing the presentation of the rating was also discussed. People are used to seeing stars as rating measures, but there is an assumption that the consumers themselves assigned or drove the star rating. It can be difficult to use stars to understand how facilities perform based on the current data that are reported on DFC (i.e., data on clinical measures). One TEP member stated that they would be happy to use stars based on consumer-reported data about facilities. Another member noted that the goal should not be to re-build the star ratings and DFC over again, but to add to the star ratings.

Another TEP member took the opportunity to highlight sources of where some of the dissonance among the current CMS DFC Five Star Ratings, methodology, and patient advocate concerns has occurred. The TEP member indicated that the originally stated CMS intent of patient- and consumer-driven rating systems—and in particular the DFC Five Star Ratings—was to provide a service to patients, based on information patients determined to be useful in the selection of their own care. The TEP member highlighted a quote from former HHS Secretary Kathleen Sebelius, who expressed similar concerns about the CMS quality rating system for Skilled Nursing Facilities, which led to a subsequent revision of that rating system. Speaking before the American Society on Aging in 2010, Secretary Sebelius stated, “We need to have some standards; they need to be clear, need to be accurately measured and if everyone ends up being excellent or everyone ends up failing, so be it. But somehow this sort of bell curve seems to have some inherent flaws.” The TEP member further quoted Secretary Sebelius as she responded to critics of the Skilled Nursing Facility rating system when she stated “The last thing we want to do is have an arbitrary bell curve just for the sake of having a system.”

Further, to again highlight the gulf between expectations created in advance for the DFC Star ratings as a patient-driven tool and the current tool, the TEP member quoted from a blog written on the CMS website on June 18, 2014 by CMS Deputy Administrator for Innovation and Quality and Chief Medical Officer Patrick Conway. In the blog post Dr. Conway stated “when buying a product or service, looking at ratings can often help narrow down the choices. Some websites offer “star” ratings that give information about the quality of the products and services they offer. Wouldn't it be helpful to have the same kind of ratings when choosing a health care provider?”

Finally, the TEP member quoted from a CMS slide presentation that was distributed by CMS to national providers on July 10, 2014 in an effort to officially explain the DFC Star Ratings. Within the slide

presentation, it stated very clearly that “Consumers are the primary audience for Compare websites, along with other important stakeholders” and, consistent with the Digital Government Strategy and the National Quality Strategy, the ratings will “report what is most important to patients in a way they can understand.”

The TEP showed interest in more comprehensive ESRD resources, and one TEP member stated interest in making educational videos for ESRD and CKD patients.

In recapping day one, a TEP member also noted that the group had discussed removing the transfusion measure. Another TEP member noted that, following the joint discussion of the prior day, they no longer supported removing the transfusion measure because of the possible adverse impact transfusions have on transplantation. Other TEP members noted that they were more interested in adding information (e.g., a new measure) to the site, not removing measures or information. They stated that they wouldn’t want to remove anything, but wish to make sure that information is available at different consumer levels. It was noted that it is important for patient education to take place so that patients understand why the measures reported on DFC are important to them. It was also noted that consumer testing of the current measures provided on the site is important.

The responsibility for clinical measure outcomes was raised by TEP members and discussed. Workgroup members were against the concept of clinical outcomes being the sole responsibility of the provider. The question was also raised about how consumers should be educated about both the facility’s and patient’s roles in measure outcomes. A TEP member responded that the more educated patients are in their own health care, the more they think it is their responsibility to have better clinical outcomes. However, the majority of patients think the clinic is responsible for what happens to them. Other TEP members agreed that the more patients are educated, the more responsibility they take for themselves. One TEP member explained that they believed that the posted facility scores reflected how well the individual patient did on meeting clinical outcomes. They stated that they also understand that it is the doctor who is responsible for bad outcomes at the clinic, but that it is also a partnership between the patient and the facility staff. They felt that the patient does also feel responsible for the clinical outcomes in addition to the doctor. It was also noted that the average patient may not look at the clinical measure scores posted at the facility, or care about the scores. Instead, they care about how they feel and their role as a patient. Others stated that the average patient may care about the facility’s quality scores but doesn’t know how to express how they care. Patients may not have the opportunity to get more information about the scores. Related discussion took place on implementation of peer-to-peer training and education at dialysis facilities as a means to disseminate information about measure scores and star rating.

4.5b DFC Website Recommendations

Suggestions for the DFC site included:

- providing a resource list;
- being able to select which items (e.g., facility data and other information) are important to the user as part of a custom rating;
- making the website easier to use;
- having website viewing options for non-English speaking or English as a second language users; and
- educating doctors and dialysis clinics about how to use the information provided on the DFC website.

The workgroup members also discussed the frequency at which people would check DFC. Many noted that it would likely occur when a patient is traveling and needs to find an alternate clinic for dialysis, or

when a patient is unhappy with their current clinic. One TEP member also mentioned that it would be useful for DFC to display information about what insurance the clinic accepted. Others noted that if it was updated annually, they may check it every year to see what changes occurred in the facility (e.g., performance changes, or if other new information is reported).

One TEP member offered recommendations on how to rollout the DFC Star ratings. They stated that there are three main audiences for the DFC website: patients, professionals, and stakeholders. Each group has a different focus. Patients primarily communicate from person to person, and would need technology that is easy to share and intuitive. Stakeholders need the source to be credible, and want to be part of the process. Stakeholders would be able to send mass emails to the patients connected to their organization to encourage them to view the DFC site. Professionals can promote the website through professional conferences such as held by ASN (American Society of Nephrologists), Kidney Week, and other events. The TEP member stated that when designing the site, stakeholders should be consulted to help beta test the website. The stakeholders would invite people to visit the site and give feedback. The TEP member recommended asking several organizations to encourage all of their viewers to go on to the site on the same day, in order to try to crash the website (in order to test the website user capacity, so there isn't a crash when it is rolled out. The TEP member recommended inviting the stakeholders to a pre-rollout briefing via conference call or in-person. During the meeting, the stakeholders would be walked through the website. Then, a briefing should be held for the media, which will help broaden the audience and inform caregivers of the website. On the day of the rollout, stakeholder groups would be asked to be engaged in sharing the content.

The TEP member recommended adding a tagline to the DFC website that says the website has patient-driven information. The TEP member recommended further promoting the patient voice on the DFC website by having videos that pop up when terms are scrolled over. The videos could feature patients explaining the meaning of a certain topic or term and why it may be important to them or impacts their care. The TEP member also recommended having features on the DFC site that can be shared on social media, which could encourage more traffic on the website. The patient's voice on the site can describe how it impacts care.

One TEP member stated that the Affordable Care Act calls for information to be transparent, easily understood, and widely available to improve quality of care and patient satisfaction. The TEP member stated that the clinical outcomes content should be open to patients and the public. The TEP member stated that in order to improve quality and satisfaction, the DFC website needs to include information that matters to the patient experience of care and reports on patient satisfaction.

One TEP member noted that patient organizations play a major role in bringing the star ratings tool to the attention of the kidney community. One TEP member stated that the patient advocacy groups had concerns about promoting the star ratings website. These groups felt it was confusing and was not patient-driven or consistent with original expectations created by CMS when the program was first presented to the renal community. The TEP member concurred with this assessment and indicated that although their organization (another large national patient organization) had a link to the five star rating tool on their website, they had not been active in promoting it to their members because of how confusing it was in the current form.

One TEP member also stated concerns about providers pushing back on the star ratings program. This was the last point that was raised before conclusion of the Public Reporting/Patient and Consumer Understanding Workgroup meeting.

4.6 Methodology Workgroup-Day 2 Discussions

4.6a Summary of Day 1 Recommendations

Addressing small facilities: the TEP is interested in getting more information about how small facilities are categorized by the star ratings. They would like to see more analyses from UM-KECC that will demonstrate if some measures are more sensitive to small facilities than others. In addition, they would like to look at the size of the facilities that fall into the tails of each measure distribution. Once they have further information, they can continue to discuss what methods, if any, need to be added to the methodology to address facility size.

Missing values: The TEP would like to evaluate two alternatives to the current imputation strategy. The first is the “nearest neighbor” strategy, and the second is to use a regression analysis to predict the score. UM-KECC will evaluate these options by conducting more analysis.

Website suggestions: the TEP recommended that CMS provide on the website more information about how a particular rating was reached. For example, design a display where the patient could adjust the rating if one particular measure is not of interest to them. Also, consider showing comparisons based on geography.

Variability and uncertainty: There is interest in including some information about variability and uncertainty in the measures, particularly for small facilities. This may be accomplished through the incorporation of confidence intervals.

Historic data: The TEP recommended that CMS include historic data in the calculations, so that patients can see trends over time. The TEP noted that small facilities may have very different results from year to year which would be smoothed if multiple years were used in the calculation. Using historic data would also decrease the ability to interpret improvement over time.

4.6b Discussion

Small Facilities

One of the TEP members stated that the exact cutoff for the definition of a “small” facility needs to be determined. In addition, the TEP should clearly define the problem with small facilities that they are trying to solve. Statistical solutions that address small patient numbers sound appropriate, but it is also important not to adjust away confounders. There are other characteristics of small facilities to consider, such as new or rural facilities.

Another TEP member explained that while one can look at characteristics of small facilities, the real issue is that the average patient count per facility is 60-70, which means many of the facilities may not have a score. This is why imputation is important. The TEP needs to identify the best analytical approach to work with the information available that will allow valid scores to be calculated. Small facilities are likely to score in the extremes, as a single observation from a patient can influence a small facility score one way or the other.

The TEP reinforced the request for analyses from UM-KECC that describe the characteristics of small facilities, and the facilities that fall into the tails of the distributions.

Incorporating New Measures

UM-KECC asked the TEP to comment on the pros and cons of incorporating patient-reported measures into the star rating, and whether it is appropriated to combine the patient-reported measures with the clinical measures. One TEP member responded that patient measures that use similar scales such as a Likert scale may be easy to combine together into their own separate domain. Another TEP member noted that the design of the domains should take into account the fact that the measures will change over time—the method of assigning/calculating domains should be stable enough to accommodate the removal or addition of new measures.

To account for variability and confidence limits, one TEP member asked if it was possible to incorporate fractions into the stars, e.g., 3.5 stars. Zagat was used as an example. Since mixing the clinical outcomes with patient-reported outcomes or quality of life measures could be challenging, perhaps having categorical stars would be an attractive way to display data. Zagat also sets specific ranges for each star (between 25 and 30 is a five star). The TEP member felt that this method was much more reflective of reality. However, some TEP members felt the stars with fractions might suggest more precision than there actually is in the star ratings.

Ranking Methods

One TEP member reiterated the question of whether the rankings should match the natural distribution of the measure, or whether it makes sense to use the probit transformation to put rankings into a “normal” distribution.

Another TEP member explained that the z-score approach stretches out the scale and re-centers it at its mean. The approach preserves the distributional shape and the tails. However, there is the possibility that the facilities that fall into the tails do not “deserve” to be there. The probit transformation was intended to mitigate that issue by shrinking the tails. A TEP member responded that it was not just the tails that is the issue; the probit ranking spreads the scores of facilities that are ranked very closely together.

Another TEP member responded that for a relative ranking goal, the probit transformation makes sense, and it actually shrinks facilities to the middle. But if the goal is not to design a relative ranking system, it might be possible to use a simpler approach. For example, one could look at the distribution of each measure and identify the cut points for each measure. Based on these cut points, one could compute stars for each measure and use these as the basis for the overall star rating. There are statistical and methodological reasons why the final score may behave better with the current more sophisticated methods. However, the simpler method might be more interpretable.

Imputation

The TEP reinforced its request to UM-KECC to investigate the possibility of incorporating more years of data into the measure calculations (to reduce the need for imputation), along with performing the simulations on the two alternative imputation methodologies. Sensitivity analyses will need to be performed on all possible options.

One TEP member noted that 18% of facilities have some imputation in their score. The use of imputation to the mean is difficult to explain to clinical staff—it is very deflating for facilities to know that a facility gets a score of 50 even in the absence of reporting. Another TEP member noted that the “nearest neighbor” imputation approach is the easiest to explain, and seems to make the most sense.

Uncertainty

UM-KECC explained that the original star rating is the point estimate based on the facility's data, but there is uncertainty around that estimate. The uncertainty of the individual measure estimates would impact the overall star rating. As a result, one needs to reexamine how each individual measure is calculated. That is, to calculate the uncertainty for each measure, and translate that into the uncertainty for the star ratings.

One TEP member pointed out that the foundation of this discussion is the issue of whether the star rating for a particular facility should be taken as an absolute number, or whether a confidence interval should be used. If the TEP is interested in the latter, it is necessary to develop a mechanism for estimating that range. UM-KECC suggested bootstrapping at the facility level, which was supported by several TEP members.

Another TEP member agreed with the idea of emphasizing the uncertainty of the component measures used in the star ratings, but the rating itself is also uncertain. In addition, they may wish to consider refining rankings by including decimals/fractions (e.g., 3.2). However, another TEP member pointed out that the use of fractions might imply precision more than uncertainty to patients trying to interpret the rating, and star ratings that use fractions would need to be carefully presented. The TEP agreed that The Public Reporting/Patient and Consumer Understanding Workgroup should be consulted on this issue.

Adding/Removing Measures

The group discussed the implications of adding or removing measures to the star ratings. This is not an immediate problem, but the discussion is relevant for the longevity of the system. They revisited the issue of whether or not to combine patient experience measures with clinical outcomes. CMS explained that there is an example of where they are treated separately: Nursing Home Compare. They calculate a rating for each domain, and then use that to construct their overall rating, which is not a pure average. The TEP agreed that the two types of measures should remain in separate domains in the star rating calculations.

In the current methodology, when measures are removed or added, factor analysis needs to be recalculated to determine the domain constructs. One TEP member was in favor of eliminating domain constructs to make it easier to add or remove measures. Another TEP member noted that with the current list of measures, removing the domains would probably have very little effect. But if there are many more measures added to the ratings in future, some may be segregated and the rating will be imbalanced without the use of domains.

Another option is to decide on domains for areas that are clinically connected—as new measures were considered, new domains would be created if necessary. This is not driven by data, but it is stable over time, and it allows for an easier interpretation. One TEP member noted that they would not be in favor of statistically defined domains, and would rather remove the concept altogether rather than forcing measures into domains based on their statistical relevance. A TEP member proposed the idea of creating weightings based on the strategic lifecycle of each measure—some measures will become less valuable over time, and can be phased out.

Another TEP member questioned the idea of pre-specified domains by relevance, wondering whether there are measures that can be grouped together. A number of TEP members stated that it was possible to create clinically relevant domains (e.g., patient safety, patient experience, etc.). For example, having a separate star rating for safety might be more helpful for patients.

One TEP member pointed out that if the domains are removed, it would change how missing data are handled. Currently, if a facility is missing a whole domain, the facility will be unrated. The number of

measures that would need to be reported in order to get a score would need to be empirically determined. CMS did note that this may ultimately be a policy decision.

Thresholds and Improvement

The group discussed the idea of setting thresholds for the measures in the star ratings. For some measures one might draw clinical lines about what is good/bad performance. In the data, there are no obvious gaps, so setting thresholds would be based on clinical judgment. However, one TEP member stated that using thresholds means commitment to fixed performance standards.

One TEP member explained further the utility of relative ranking. If nationally all facilities move into a clinically acceptable high level, their rankings will continue to indicate a range of performance. It would be helpful to associate the absolute level with each facility so that one can clearly view what, for example, a three star rating means.

Another TEP member felt that for some of these measures, if a facility meets a certain threshold, that shows optimized care and deserves to be rewarded. Additionally, thresholds acknowledge the idea that the measures such as those for vascular access are not always appropriate for all patients. Not every facility should reach 100% success on all measures. UM-KECC suggested that in such cases, a mixed measure could set a threshold of lower than 100% for “full credit” while still using a continuous ranking for those below the threshold. For example, a 60% threshold for fistula use would not differentiate among facilities at or above 60%, but would differentiate among facilities falling below 60%.

One TEP member noted that if one observes the mean size of the facility and some of the clustering population performance, the difference between 1-and 5-star facilities can seem artificial. Another TEP member replied that using a threshold would not resolve this issue.

CMS asked the TEP to comment on the appropriate thresholds for each measure. Using the SMR as an example, if CMS is setting an “acceptable” rate of death, this could be very problematic. This seems to be the largest stumbling block, which cannot be totally resolved. One TEP member noted that thresholds could be based on the literature, or developed by an advisory group.

A TEP member stated that methodological issues would arise when combining measures that have a threshold with those that do not. However, in general, there would be statistical means to incorporate thresholds/improvement into the ratings. Some TEP members expressed that improvement might be difficult to estimate with the noisy data.

The group then touched on improvement. CMS explained that improvement is incorporated into the Nursing Home Compare star ratings, and the issue that arose was measure creep in terms of performance. They began with a particular pre-set distribution, retained those thresholds, and followed star ratings through the course of several years. Eventually, the vast majority of facilities were four and five star. To eliminate this issue, annual rebasing may be necessary (similar to the QIP). If the measure performance hits a ceiling where most facilities are performing well, it can be considered for retirement.

4.7 Summary of the Final Joint Workgroup Discussion Session (Day Two)

4.7a Public Reporting/Patient and Consumer Understanding Workgroup Recommendations

The Public Reporting/Patient and Consumer Understanding Workgroup presented their summary and recommendations at the final session of the TEP (see Appendix G).

The workgroup's general summary highlighted the following issues.

1. For risk-adjusted measures (SMR, SHR, STeR), SES adjustment is an important factor. Some were in favor of including such adjustments, while others feel that the potential benefits and drawbacks need further exploration.
2. The workgroup stated that the methodology used for the star ratings is difficult to understand for intended users (patients and other consumers), and not consistent with other online public rating systems. DFC language needs to more clearly explain the difference between the DFC Star Ratings reported and other consumer sites such as Yelp, and that the Compare sites are different from other consumer sites that use stars which are based on direct consumer reported satisfaction.
3. The Public Reporting/Patient and Consumer Understanding TEP Workgroup stated that CMS needs to decide if the site is intended to assist patients in decision making or if it is intended as a quality improvement effort. Depending on the intended audience, the presentation of the website may be very different. In general, as currently presented, the overall perceived relevance of DFC measures and star ratings for patients/consumers is low.¹

The Public Reporting/Patient and Consumer Understanding Workgroup Presented the Following Recommendations:

1. Setting an established standard to assess performance in the star ratings is preferred over relative or "curved" rankings when possible.
2. Include multiple levels (gradations) of standards or thresholds to indicate partial achievement of a standard.
3. Include the opportunity for a facility to provide comments/explanations of its star rating.
4. Greater consistency across ESRD programs- some members were clearly more supportive of this recommendation and others less certain.
5. Consensus for addition of new measures: members strongly felt additional measures are needed and that retiring measures was not a priority.
6. Adding measures on safety outcomes (e.g., infection outcomes, falls, medication errors, adverse event reporting, etc.).²
7. Adding measures of patient-reported outcomes (e.g., quality of life, patient-assessed quality of care).
8. Adding measures of facility staff (e.g., assessment of staff training/performance, promoting modality choice, staff responsiveness for patient concerns, adequate staffing).

¹ The Public Reporting/Patient and Consumer Understanding Workgroup discussion included that the workgroup thought clinical outcomes were important, but that patients are also interested in the patient reported ratings on subjects identified by patients as important.

² Please see section 4.2f (Discussion of Relevant DFC Measures) for more measure recommendations.

9. Broader consumer testing of current measures in the star rating program is needed to assess relevance to consumers (examples given included SMR, SHR, vascular access, and dialysis adequacy).³
10. When asked about possible measure retirement, the patient workgroup identified hypercalcemia as the most likely candidate for removal. The workgroup was more concerned with development of additional patient-reported metrics than with removal of current measures.

Star Ratings Rollout Recommendations

The Public Reporting/Patient and Consumer Understanding Workgroup stressed the importance of collaboration with CMS, patients, and stakeholders in order to effectively rollout the star ratings. The patient and consumer workgroup stated their recommendations for the star ratings rollout. They mentioned the importance of involving patients, stakeholders, and professionals in the star ratings process of communicating with and educating the community. Patients and patient organizations are in a position to disseminate information about the star ratings to other patients through in-person interactions and social media. Stakeholders have the opportunity to promote the star ratings through a collaborative kickoff, emails, and web links from their websites. Professionals and professional organizations have the opportunity to promote through ASN Kidney Week, and other events.

The workgroup further emphasized the importance of working with stakeholders and receiving stakeholder feedback in the star rating development process. The workgroup presented a rollout model that is similar to what has been used for the rollout of other federal programs. Initially, the star ratings should receive feedback from stakeholders. After incorporating feedback, beta testing should be conducted with patients. Then, a pre-rollout briefing should be conducted with stakeholders. Afterwards, an embargoed brief should be released in which the media would be informed about the star ratings. Finally, a full rollout should be conducted jointly with the stakeholders.

4.7b Methodology Workgroup Recommendations

The Methodology Workgroup presented the major discussion points of scoring, weighting, categorization, missing data (imputation), uncertainty, and data presentation. The workgroup then presented their recommendations on the discussion points (see Appendix H).

Discussion of the major points began with scoring. The first issue is how to combine measures that have different numeric scales and distributions. The second issue is how to reduce the effects of imprecise measures (e.g., for small facilities). The current method is to rank measures using probit scoring. The advantages of the probit scoring method are that it has strong statistical properties, addresses the previous two issues, and it maintains relative standings in terms of the scores. The disadvantages are that it does not retain information on actual distance between values in the original scale of measurement and that it may appear less clinically transparent.

An alternative scoring method is z-scores, a simple change of original units to standard deviation units away from the mean value. The advantages are that it retains information on the original spacing, and it can combine measures that are on different scales. The disadvantages are that it does not address how to reduce the effects of imprecise measures (e.g., wide variance of measure values in small facilities). Another disadvantage is that this approach gives more weight to extreme values when combining measures.

³ The Public Reporting/Patient and Consumer Understanding Workgroup expressed interest in adding one or more domains to the current star rating that are based on patient experience of care.

The goal of measure weighting is to combine measures into a single measure that can be transformed into stars. The first stage is to combine measures that are effectively assessing the same thing (e.g., through use of factor analysis). Factor analysis avoids letting one characteristic of facility quality dominate the overall ratings because it has multiple contributing measures. The factor analysis must be re-calculated when measures are added or removed. Data-driven methods such as factor analysis may lack direct clinical interpretability. Instead, clinically determined groupings/domains could be used. It was noted the seven current measures are effectively equally weighted, so this step of using clinical determination is not a central part of the current final scoring. However, this could change with additional or fewer measures. The second stage is combining all of the domains into a single summary score. Currently measures are averaged within domains and domains are equally weighted and combined into a final score.

Discussants also summarized points about how star rating categories were determined. Currently the stars are defined based on fixed percentiles. Facilities are ranked highest to lowest based on final scores, then binned into star rating categories. The top ten percent receive five stars. The next highest twenty percent (approximately 70-90%) receive four stars. The middle forty percent receive three stars (approximately 30-70%). The next lowest twenty percent (approximately 10-30%) receive two stars. The bottom ten percent receives one star. This method essentially is “grading on a curve” and is methodologically appropriate for the goal of relative ratings. It was noted that three stars represent the national average, and other categories are relative to that. The chairs of the methodology workgroup summarized the conclusion of the Methodology workgroup that the methodology developed was valid and had strong statistical properties.

The Methodology Workgroup presented the following recommendations:

1. Anchor the stars in clinically meaningful terms.
 - a. Average scores for each clinical measure and the combined domains can be reported numerically and/or graphically for each star rating category.
 - b. Actual facility-level measures and associated percentiles, measures of uncertainty, etc. should also be reported on DFC.
2. Impute missing values in a more informative way.
 - a. Current approach: no final score is assigned to a facility if all measures in a domain are missing; or, if there is at least one non-missing measure in a domain, a national average is assumed for the measures in that domain that have missing values.
 - b. The limitation of current approach is that the average facility size is around 60-70 patients. Small facilities are more likely to have missing values and less likely to be scored.
 - c. Recommendation: use more facility level information to impute.
3. Need to present information on uncertainty in ratings.
 - a. Proposed several possible approaches, which KECC will investigate for performance properties.⁴

4.7c Final Joint Session Discussions

The patient group reiterated the opinion that the methodology behind the star ratings should reflect a rating that is most useful for patients, which would potentially involve moving away from relative ranking toward absolute thresholds. Members of the methodology group acknowledged the patient group concerns, and explained that the methods can accommodate ratings based on absolute thresholds or a

⁴ The recommendations for investigating uncertainty are in section 4.3c and 4.6b.

combination of relative rankings and thresholds—they are looking for guidance from the patients on what is the most useful.

One TEP member asked the methodology group to clarify how they would incorporate patient-reported outcomes into the star rating methodology. Would measures of patient-reported outcomes (e.g., quality of life or patient experience of care, satisfaction) be of equal weight? Would they be part of their own domain? A TEP member from the methodology group replied that there are a number of different ways to incorporate these patient-reported outcome measures, and a rating can be created based on the patient reported outcomes, which can be done by applying the current UM-KECC method.

One TEP member stated that the Patient Protection and Affordable Care Act calls for information to be transparent, easily understood, widely available, and to improve quality and satisfaction. The TEP member agreed that the clinical outcomes content should be open to patients and the public. The TEP member stated that in order to improve quality and satisfaction, Five Star Ratings need to include information that matters to patients on the patient experience of care and reports on patient satisfaction.

A TEP member sought clarification on the Methodology Workgroup's recommendation to anchor the stars in clinically meaningful terms. A TEP member from the Methodology Workgroup explained that in addition to the measure results for individual facilities, the DFC website should incorporate more information about what each star rating means in general. For example, when one looks at a three star facility, one should also be able to see and compare the range of fistula percentage (the average and variation) for other three star facilities. This is essentially a profile of average performance for each of the star categories, making the star ratings more transparent and interpretable.

One TEP member asked Dr. Andress (CMS) if the star rating methodology will be different when the star ratings are released in 2016. Dr. Andress stated he suspects there may be changes to the star ratings, but that he is not able to make a commitment until having more information on what the changes will entail. The TEP member recommended that an earlier release of an outline of changes would be helpful. Dr. Andress stated that he would be in contact with the TEP member soon to discuss the outline for potential changes.

4.8 Public Comment Period

A thirty-minute public comment period was held at the conclusion of the In-Person TEP Meeting on April 28, 2015. At the beginning of the public comment period, two individuals indicated they would be making comments. One public comment was provided by Kathy Lester, JD, of Kidney Care Partners (KCP). The second commenter later waived his opportunity to comment.

Kathy Lester: "First of all, I am Kathy Lester. I am here on behalf of Kidney Care Partners (KCP). I think many of you know the organization. It is about thirty-two different members of the kidney care community. It includes physicians, nurses, healthcare professionals, patients, patient advocacy groups, manufacturers, and of course the dialysis facilities large to small.

I appreciate the difficulty in pulling consensus together given the work that we do with that group. First, I would just like to thank CMS and UM-KECC for probably the most transparent TEP that we've been part of. This has been something you all know that is important to us. And very grateful for the opportunity to have people to call in, to be able to sit in a room, be in workgroup meetings, and participate.

KCP has always been supportive of publicly reported quality programs. We were the first at the table and actually asked Congress to create the ESRD Quality Incentive Program (QIP), which was the first value-based purchasing program in Medicare. And as you saw from our comments on Five Star, we are very supportive of the concept of trying to have an easy to understand set of quality metrics for patients. I think, however, as you heard from many around the table and some of our comments- we do have two major concerns that I think the TEP has discussed and look forward to seeing what comes out of it.

The first is that we believe the program should rely on transparent performance benchmarks rather than a relative ranking. We also believe that it is essential to use valid measures that are important to patients and that facilities actually can control or are actionable for those facilities. The patient workgroup members really echoed the concerns unanimously in their discussion about the methodology and the need for performance standards and we were very grateful for that. We were pleased that the methods workgroup was able to discuss that as well and to hear that there is a methodology possible for doing that. In a system that is designed to make it easier for patients and their families to understand, to evaluate, and to choose quality dialysis providers, quality performance really needs to be translated in a way that is intuitive for them. Consumers today certainly would not guess today that a three star facility for example would be in the top 31% of facilities national wide. So we do think there is work to be done here.

I did want to address the comment of whether benchmarks could be used. They actually can be. There should be no issue in implementing that within the system. Other Medicare star rating program uses benchmarks and in particular the Quality Incentive Program for ESRD has those benchmarks. There are good models out there to look to. To be clear I think using performance benchmarks rather than a relative ranking does not mean that the benchmarks never change. You see that in these other Medicare programs the way that they do evolve over time. And there are clear methodologies for using those.

What it does mean is that CMS would not be using a methodology- that as the methodologists pointed out- that do exaggerate the spatial differences between facilities that may not have any clinical, meaningfulness for patients. Similarly, we think that measures in such a system should be linked to actionable items that are within the dialysis facilities' control so patients know what is going to happen to them in those facilities. Metrics like mortality and hospitalization are very important but issues such as car accidents, or cancer admissions, or mortalities because of cancer should not be part of those metrics. So it is refining I think what we have there. In sum, KCP supports the recommendations from the patients to adopt the performance benchmark methodology for Five Star.

We also emphasize that CMS should ensure that the metrics used are measurable and actionable for the dialysis facilities directly. In this way, Five Star would be a useful program and provide patients and their families with information about actual facility performance. In conclusion, we would really welcome the opportunity to continue the open dialogue and appreciate all the work and the time that people have devoted here today and are offering to help in any way that would be most appropriate. Thank you."

5. Potential Follow-up TEP Teleconference Calls

Potential follow-up TEP teleconference calls may be determined necessary at a later time. At this point no calls have been scheduled.

6. Appendices

Appendix A: Technical Expert Panel Charter

Project Title:

End Stage Renal Disease Dialysis Facility Compare (DFC) Star Ratings Technical Expert Panel (TEP)

Dates:

February – August, 2015

Project Overview: The Centers for Medicare & Medicaid Services (CMS) has contracted with The University of Michigan Kidney Epidemiology and Cost Center (UM-KECC) to review the methodology developed to produce the DFC Star Ratings. The contract name is the ESRD Quality Measure Development, Maintenance, and Support contract. The contract number is HHSM-500-2013-13017I.

Starting Jan 22, 2015, the Medicare DFC website displays star ratings of dialysis facilities. The star ratings give a broad idea of the quality of care patients with End-Stage Renal Disease (ESRD) and their families can expect from dialysis facilities, and provides the ability to compare facilities to each other. The introduction of these ratings is part of CMS's plan for all of the Medicare Compare sites to make quality information more accessible to patients, and other key audiences, including providers and policymakers. The goal is to increase knowledge, understanding, usability, and satisfaction with the Compare sites.

UM-KECC uses CMS's structured and standardized approach to measure development, including steps to ensure input from experts and the public. As part of this effort, UM-KECC is seeking input from individuals with relevant experience and expertise who can provide a variety of perspectives on the methodology behind the star ratings. This group will serve as the rating's Technical Expert Panel (TEP).

The TEP is a group of subject matter experts and stakeholders who provide input on the development and maintenance of measures for which the contractor is responsible. For this project, TEP members will review the current methodology of the DFC star ratings, and provide feedback on how the methodology could be improved. They will also have the opportunity to provide feedback on how the star ratings are conveyed to the community on the DFC website.

The TEP will represent a diversity of perspectives and backgrounds. Members will be selected for the input they can provide based on their personal experience and training or organizational perspective. Given that the audience for the star ratings is primarily patients; this TEP will have heavy representation from patients and patient advocates.

We anticipate that participants will be divided into two workgroups to ensure that all participants can contribute based on their expertise and experiences. This structure will also maintain an optimal smaller group size to allow for greater in-depth and focused discussion. We anticipate that the meeting will take place over two days and will include an opening and closing session where the entire group meets to discuss the overarching themes and agenda of the TEP as well as understand the specific issue areas each respective workgroup will be discussing. The primary focus of workgroup one will be on the star rating

methodology; the focus of workgroup two will be to review and discuss public reporting, and patient/consumer needs and goals. The topics discussed by the workgroups are not expected to be mutually exclusive, but will differ in the type of conceptual and technical detail devoted to issues of methodology, and public reporting and patient/consumer understanding. When completing the letter of interest please indicate in which workgroup(s) you would be willing to participate (Methodology or Public Reporting/Patient and Consumer understanding).

Project Objectives:

The TEP will evaluate and make recommendations on the DFC star rating methodology and display. Specific objectives include:

1. Review of the statistical methodology behind the star rating calculations
2. Review of the measures used in the star ratings
 - a. Consider measures for retirement
 - b. Consider measures for future implementation
3. Review the readability and presentation of the star ratings on the DFC website

TEP Objectives:

- Review the star rating methodology that is currently used to report star ratings on DFC.
- Provide input on measure selections for the domains, domain-level star scoring methodology, and overall star ratings methodology.
- Make recommendations regarding measures to be added or removed, and the star ratings methodology.

Scope of Responsibilities:

The role of the TEP and each member is to provide advisory input to UM-KECC regarding the star ratings system.

Role of UM-KECC: As the CMS measure development contractor, UM-KECC has a responsibility to support the development of quality measures and public reporting for ESRD patients. The UM-KECC moderators will work with the TEP chair(s) to ensure the panel discussions are focused; during discussions, UM-KECC moderators may advise the TEP and chair(s) on the needs and requirements of the CMS contract and the timeline, and may provide specific guidance and criteria that must be met with respect to CMS requirements.

Role of TEP chair(s): Prior to the in-person TEP meeting, one or two TEP members are designated as the chair(s) by UM-KECC and CMS. The TEP chair(s) are responsible, in partnership with the moderator, for directing the TEP to meet the objectives of the TEP, including provision of advice to the contractor regarding the star rating system.

Duties and Role of TEP members: As defined by CMS in the Measure Management System Blueprint, TEPs are advisory to the measure contractor. In this advisory role, the primary duty of the TEP is to review the existing star rating methodology and supporting materials, and provide feedback to UM-KECC.

March and April 2015 attend pre-TEP conference calls as necessary. Attend one in-person meeting in April of 2015 (dates are yet to be determined) in Baltimore, MD, and be available for additional follow-up teleconferences and correspondence as needed.

The TEP will review, edit (if necessary), and adopt a final charter at the first teleconference. The discussion will be on the overall tasks and goals/objectives of the TEP.

During the In-Person Meeting: The TEP will review the existing star rating system. The key deliverables of the TEP in-person meeting include a summary report documenting the discussions and decisions that take place during the In-Person Meeting. The report will outline preliminary TEP recommendations.

At the end of the two day meeting the TEP chair(s) and TEP members will prepare a summary of recommendations. As necessary, the TEP chair(s) will have additional contact with UM-KECC moderators to work through further discussion of proposed recommendations. After the In-Person Meeting (approximately April –August, 2015): TEP members will review a summary report of the TEP meeting discussions, recommendations, and other necessary documentation forms.

Guiding Principles:

Potential TEP members must be aware that:

- Participation on the Technical Expert Panel is voluntary.
- Input will be recorded in the meeting minutes.
- Proceedings of the in-person meeting will be summarized in a report that is disclosed to the general public.
- Potential patient participants may keep their names confidential, if they wish to do so.
- If a TEP member has chosen to disclose private, personal data, that material and those communications are not covered by patient-provider confidentiality.
- All questions about confidentiality will be answered by the TEP organizers.
- All potential TEP members must disclose any current and past activities that may pose a potential conflict of interest for performing the tasks required of the TEP.
- All potential TEP members must commit to the expected time frame outlined for the TEP.
- All issues included in the TEP summary report will be voted on by the TEP members.
- Counts of the votes and written opinions of the TEP members will be included, if requested.

Estimated Number and Frequency of Meetings:

- TEP members should expect to come together for one to three teleconference calls prior to the in-person meeting held April 2015, in Baltimore, MD.
- The in-person meeting April 2015 (final dates to be determined).
- After the in-person meeting, additional conference calls may be needed.
- Additional workgroup sessions as determined

Date Approved by TEP:

TBD

TEP Membership:

TBD

Appendix B: Public Reporting/Patient and Consumer Understanding Pre-TEP Conference Call Minutes

**ESRD Quality Measure Development, Maintenance, and Support Project
End Stage Renal Disease Dialysis Facility Compare (DFC) Star Ratings
Public Reporting/Patient and Consumer Understanding Pre-TEP Conference Call #1
April 13, 2015 4:00 – 5:00pm**

| TEP Members | UM-KECC | CMS |
|---|--------------------------|-----------------------|
| Paul Conway | Joseph Messana, MD | Joel Andress, PhD |
| Joanna Galeas | Yi Li, PhD | Elena Balovlenkov, RN |
| Joseph Karan | Claudia Dahlerus, PhD MA | |
| John (Jack) Reynolds | Rich Hirth, PhD | |
| Tonya Saffer, MPH | Jordan Affholter | |
| Chris Sarfaty, MSW | Casey Parrotte | |
| David White | | |
| “Anonymous Patient A” (withheld by request) | | |

Introductions

Dr. Joseph Messana welcomed everyone to the Pre-TEP conference call, and thanked the TEP members for their time. Dr. Messana introduced himself as a clinical Nephrologist working for University of Michigan Kidney Epidemiology and Cost Center (UM-KECC). Dr. Messana introduced Dr. Yi Li as the Principal Investigator for the ESRD Quality Measure Development, Maintenance, and Support Project contract for the Center for Medicare/Medicaid Services (CMS). Dr. Messana stated that the University of Michigan Kidney Epidemiology and Cost Center (UM-KECC) is the CMS contractor for developing ESRD quality measures.

The Star Rating TEP has two workgroups: Methodology and Public Reporting/Patient and Consumer Understanding. Dr. Messana will be leading the Patient/Public Reporting Workgroup and Dr. Li will be leading the Methodology Workgroup. Dr. Messana mentioned that Dr. Richard Hirth, Dr. Claudia Dahlerus, Dr. Ji Zhu, and other UM-KECC support staff have participated in preparations for this TEP. He mentioned that Jordan Affholter would be the contact person for the TEP members. The contact email and phone number were provided to the TEP members if they have any questions.

Dr. Messana asked the TEP Members to give a brief introduction. He also stated that one patient did request confidentiality. Patients TEP members have the option to maintain confidentiality as outlined in the blueprint and TEP charter. The confidential patient will be referred to as “Anonymous Patient A”.

TEP member Introductions

Paul Conway: Paul Conway is the President of AAKP (American Association of Kidney Patients). He has extensive administrative experience in both federal and state government. In addition, he has experience in advocacy positions and has personal experience with ESRD.

Joanna Galeas: Joanna Galeas has personal experience with ESRD.

Joseph Karan: Joseph Karan has personal experience with ESRD.

Jack Reynolds: Jack Reynolds has personal experience with ESRD and is the Vice-President of Dialysis Patient Citizens.

Tonya Saffer: Tonya Saffer is the Senior Health Policy Director for National Kidney Foundation (NKF) with work with patients on advocacy and education issues. She has familial experience with ESRD.

Chris Sarfaty: Chris Sarfaty is a Patient-Centered Collaborative Care Coach with a Master's in Social Work. She has worked in dialysis for twenty five years and has familial experience with ESRD.

David White: David has personal experience with ESRD. He is the Chair of the Patient Advisory Committee chair and a Medical Review Board member for the Mid-Atlantic Renal Coalition (MARC).

"Anonymous Patient A": She has personal experience with ESRD who choose to maintain confidentiality. "Anonymous Patient A" is a Health Care Research worker at a health care company that works under contract with CMS (non-ESRD topic areas). Her company has submitted a bid for an ESRD related contact.

The TEP members will disclose their perceived or real conflict of interest at the TEP introductions on day one of the in-person TEP meeting.

CMS introductions

Elena Balovlenkov of CMS is a nurse and the DFC lead for Public Reporting. She has been working in ESRD for 38 years and has experience working with advocacy groups. She thanked everyone for their participation.

Dr. Joel Andress is the CMS Contracting Officer's Representative for the ESRD Quality Measure and Public Reporting contract. Dr. Andress thanked all of the TEP members for volunteering their time to this project.

Dr. Messina gave brief introductions for the Methodology Workgroup Members, and noted that the Methodology teleconference will be held on April 14, 2015. The Methodology workgroup members are Dr. Richard Cook, Dr. J. Richard Landis, Dr. Franklin Maddux, Dr. Allen Nissenson, Dr. Jane Pendergast, Dr. Dylan Small, Dr. Nicole Stankus, Dr. Catherine Sugar, and Sumi Sun.

Dialysis Facility Compare (DFC)

Dr. Messina gave a brief overview of the Dialysis Facility Compare (DFC) website. Dialysis Facility Compare provides information about Medicare-certified dialysis facilities. The website also compares the services and the quality of care that facilities provide.

The reported quality measures on Dialysis Facility Compare has expanded to include thirteen reported measures (in 2015) on the website including SMR, SHR, STrR, SRR, Hgb >12.0 g/dL, Hgb <10.0 g/dL, Fistula, Catheter, Adult HD Kt/V, Adult PD Kt/V, Pediatric HD Kt/V, Hypercalcemia, and Serum Phosphorus. The SRR, Hgb >12.0 g/dL, Hgb <10.0 g/dL, and Serum Phosphorus measures were not included in the calculations for the January 2015 Star Ratings release. In January 2013, the DFC website underwent extensive revisions to include both table and graphical displays for all measures. More information including ratios and confidence intervals were added for standardized measures.

Dr. Messina used the DFC website to show the variety of Star Ratings for the three University of Michigan facilities (one 1-star, one 3-star, and one 5-star). Facilities are given a star rating between one and five stars. A national call for increased transparency and wider use of publicly reported data on health care quality was the driving force behind the Star Rating development. The DFC Star Ratings were announced in mid-2014 prior to the Star Ratings being previewed to facilities. The Star Ratings were made available to the public on DFC in January 2015. Star Ratings are calculated using most of the current Dialysis Facility Compare quality measures (listed in the previous paragraph).

Dr. Messina reviewed a few of the community concerns. He noted that commenters have raised concerns about the methodology behind the Star Ratings, such as the distribution of the final score, appropriateness of the rating categories, risk adjustments that have been used, and consistency in ranking across federal programs. He also mentioned the issue of whether it is appropriate to use relative ranking or fixed threshold when calculating the ratings. In addition, there have been comments on unintended consequences of the Star Ratings. TEP participants will be asked to discuss these questions and concerns at the in-person meeting.

Dr. Messina gave an overview of the Star Rating TEP structure and operational paradigm. The Technical Expert Panel (TEP) will take place over two days in Baltimore, MD on April 27th and April 28th. The Star Rating TEP has two workgroups: Methodology and Public Reporting/Patient and Consumer Understanding. The TEP workgroups will have joint sessions where both workgroups will be together for the discussion. The TEP workgroups will have also concurrent sessions where the two workgroups will work on the areas of their focus. The TEP meeting is open to the public and will be recorded. Patient confidentiality will be respected. There will be also be time for public comments.

The TEP Charter is the general roadmap for the TEP. UM-KECC assumes that the TEP Members have agreed that the TEP Charter is the general roadmap when by agreeing to participate. If the TEP members have questions or concerns about the TEP Charter, they should contact UM-KECC. The TEP members will release their perceived or real conflict of interest at the TEP introductions on day one of the TEP. Individual workgroup chairs will be appointed for the TEP patient workgroup.

Roles

Dr. Messina explained that the TEP serves an advisory role to UM-KECC. They are charged with reviewing

the current methodology and making recommendations on how to improve the methodology and reporting of the Star Ratings. TEP members will be expected to share their expert opinions and experience. Their responsibility is to be a respectful team member in order for the opinions of all TEP members to be heard. Also, TEP members are responsible for being active and engaged in the TEP discussions.

UM-KECC will serve as the moderator and facilitator of the TEP, with the goal of keeping the discussion focused and providing specific details about the Star Rating methodology and measures. UM-KECC is responsible for listening carefully and recording the TEP opinions and recommendations. They will also prepare a summary report after the TEP meeting, based on notes and audio recordings from the meeting. UM-KECC will make recommendations to CMS based on the input received from the TEP.

C-Tools Explanation

- UM-KECC created a web portal through the University of Michigan website, to share a variety of tools that the TEP members may find useful, such as:
 - TEP Charter and Consolidated Questions Document
 - Star Ratings materials and full pdf versions of the articles used in the supporting literature.
- Instructions for obtaining an account were circulated to the TEP members in advance of the teleconference call, and Jordan Affholter gave a brief presentation on gaining access to the C-Tools site.
- C-Tools will be the main vehicle for providing documents to the TEP. TEP members were encouraged to send materials to UM-KECC if they would like to post any materials to the C-Tools site. UM-KECC encouraged TEP members to create their accounts and let UM-KECC know if any issues arise.

Public Comments

There were no public comments received during this TEP.

Wrap Up

Dr. Messana thanked the TEP members and general public for being involved in the call.

Appendix C: Methodology Workgroup Pre-TEP Conference Call Minutes

**ESRD Quality Measure Development, Maintenance, and Support Project
End Stage Renal Disease Dialysis Facility Compare (DFC) Star Ratings
Methodology Workgroup Pre-TEP Conference Call #1
April 14, 2015 4:00 – 5:00pm**

| TEP Members | UM-KECC | CMS |
|-----------------------------------|--------------------------|-----------------------|
| Richard Cook, PhD, MS | Yi Li, PhD | Joel Andress, PhD |
| J. Richard (Dick) Landis, PhD, MS | Claudia Dahlerus, PhD MA | Elena Balovlenkov, RN |
| Allen Nissenson, MD, FACP | Rich Hirth, PhD | |
| Jane Pendergast, PhD | Jordan Affholter | |
| Dylan Small, PhD | Casey Parrotte | |
| Nicole Stankus, MD, MSc | | |
| Catherine Sugar, PhD, MS | | |
| Sumi Sun, MPH | | |

Introductions

Dr. Yi Li welcomed everyone to the Pre-TEP conference call, and thanked the TEP members for their time. Dr. Li gave a brief overview of what was going to be discussed on the conference call including introductions, the role of the TEP, and the C-Tools website. The call was open to the public and the last five minutes were set aside for public comments.

Dr. Li stated that the University of Michigan Kidney Epidemiology and Cost Center (UM-KECC) is the CMS contractor for developing ESRD quality measures. Dr. Li introduced himself as the Principal Investigator on this project and a Professor of Biostatistics at School of Public Health. He introduced Dr. Joseph Messana as a clinical Nephrologist working for UM-KECC. Dr. Li introduced Dr. Richard Hirth, Professor of Health Management and Policy at the University of Michigan. Dr. Li introduced Dr. Claudia Dahlerus, a Principal Research Scientist at University of Michigan. Dr. Li introduced Dr. Ji Zhu, a Professor of Statistics at the University of Michigan. Dr. Li also listed rest of the UM-KECC team: Casey Parrotte, Karen Wisniewski, Christopher Harvey, Natalie Scholz, Cindy Liao, Jennifer Sardone, and Zezhi (Zac) Zhang. He mentioned that Jordan Affholter would be the contact person for the TEP members. The contact email and phone number were provided to the TEP members if they have any questions.

Elena Balovlenkov of CMS is the DFC lead for Public Reporting and has been working in ESRD for 38 years. She welcomed and thanked everyone for their time and commitment. Dr. Joel Andress of CMS was also in attendance. Dr. Andress is the CMS Contracting Officer's Representative for the ESRD Quality Measure and Public Reporting contract. Dr. Andress thanked all of the TEP members for volunteering their time to this project.

Dr. Li thanked the group for making time for the commitment for attending the Pre-TEP call and committing to attending the TEP in Baltimore, MD. The Star Rating TEP has two workgroups: Methodology and Public Reporting/Patient and Consumer Understanding. The Methodology workgroup is comprised of nine experts, and Dr. Jane Pendergast and Dr. Catherine Sugar have agreed to serve as co-chairs for the Methodology workgroup.

Each TEP member gave a brief introduction.

Dr. Jane Pendergast: Dr. Jane Pendergast is a Professor of Biostatistics at Duke University. She worked at University of Iowa in the past working on an economics project related to ESRD. She also has familial experience with ESRD.

Dr. Catherine Sugar: Dr. Catherine Sugar is an Associate Professor (Statistics, Biostatistics, and Psychology) at UCLA (University of Southern California, Los Angeles). She has past experience working on Biostatics projects dealing with Nephrology. She has experience with health care analysis, scale development, and methodological techniques relating to large data sets.

Dr. Richard Cook: Dr. Richard Cook is a Professor of Statistics at the University of Waterloo, Canada. He works in Biostatistics and has past experience working in a Nephrology project.

Dr. J. Richard Landis: Dr. J. Richard Landis is a Professor of Biostatistics and Statistics at the University of Pennsylvania Medical School. He has worked on collaborative projects with investigators in Nephrology at University of Pennsylvania. He is currently on clinical research studies relating to dialysis patients. He worked at University of Michigan in the past related to large health care data sets.

Dr. Allen Nissenson: Dr. Allen Nissenson is a clinical Nephrologist and the Chief Medical Officer of DaVita Healthcare Partners. He is also a Professor Emeritus of the David Geffen School of Medicine at UCLA (University of Southern California, Los Angeles). He has experience in developing measures for kidney disease. He is a co-chair of the Kidney Care Quality Initiative, which develops kidney metrics.

Dr. Dylan Small: Dr. Dylan Small is an Associate Professor for the Department of Statistics at the University of Pennsylvania. His work is related to observational research for large databases.

Dr. Nicole Stankus: Dr. Nicole Stankus is a clinical Nephrologist and Medical Director of a dialysis facility in Chicago, Ill. She is also an Associate Professor of Medicine at the University of Chicago.

Sumi Sun: Sumi Sun is the Director of Applied Research and Data Analysis at Satellite Healthcare, which is a non-profit dialysis provider focused in California. She is an epidemiologist with a background in public health.

Dr. Li gave a brief introduction for Dr. Franklin Maddux, who was not able to attend the call. Dr. Franklin Maddux is the Executive Vice President for Clinical and Scientific Affairs at Fresenius Medical Care.

Dr. Li then gave brief introductions for the Public Reporting/Patient and Consumer Understanding Workgroup Members who had their teleconference the day before on April 13, 2015. The Public Reporting/Patient and Consumer Understanding Workgroup Members are Paul Conway, Joanna Galeas, Joseph Karan, John (Jack) Reynolds, Tonya Saffer, Chris Sarfaty, David White, and "Anonymous Patient A" (withheld by request).

Dialysis Facility Compare (DFC)

Dr. Li gave a brief overview of the Dialysis Facility Compare (DFC) website. Dialysis Facility Compare provides information about Medicare-certified dialysis facilities. The website also compares the services and the quality of care that facilities provide.

The reported quality measures on Dialysis Facility Compare has expanded to include thirteen reported measures (in 2015) on the website including SMR, SHR, STrR, SRR, Hgb >12.0 g/dL, Hgb <10.0 g/dL, Fistula, Catheter, Adult HD Kt/V, Adult PD Kt/V, Pediatric HD Kt/V, Hypercalcemia, and Serum Phosphorus. The

SRR, Hgb >12.0 g/dL, Hgb <10.0 g/dL, and Serum Phosphorus measures were not included in the calculations for the January 2015 Star Ratings release.

Facilities are given a star rating between one and five stars. A national call for increased transparency and wider use of publicly reported data on health care quality was the driving force behind the Star Rating development. The DFC Star Ratings were announced in mid-2014 prior to the Star Ratings being previewed to facilities. The Star Ratings were made available to the public on DFC in January 2015. Star Ratings are calculated using most of the current Dialysis Facility Compare quality measures (listed in the previous paragraph).

CMS and UM-KECC have made effort to address comments and criticisms of the Star Ratings. Dr. Li reviewed a few of the community concerns, including the distribution of the final score and how to categorize the final score. He mentioned the issue of whether to use relative ranking or fixed threshold. Dr. Li explained that these questions and concerns would be discussed in detail at the in-person meeting. Many of the issues have been addressed in the Consolidated Questions and Response Document on the C-Tools website and Dr. Yi Li recommended that the TEP members review this document.

Dr. Li gave an overview of the Star Rating TEP Structure and Operational Paradigm. The Technical Expert Panel (TEP) will take place over two days in Baltimore, MD on April 27th and April 28th. The Star Rating TEP has two workgroups: Methodology and Public Reporting/Patient and Consumer Understanding. The TEP will have joint sessions where both workgroups will be together for the discussion. The TEP will have also concurrent sessions where the two workgroups will work on the areas of their focus. The TEP meeting is open to the public and will be recorded. Patient Confidentiality will be respected.

The TEP Charter is the general roadmap for the TEP. UM-KECC assumes that the TEP Members have agreed that the TEP Charter is the general roadmap when by agreeing to participate. If the TEP members have questions or concerns about the TEP Charter, they should contact UM-KECC. The TEP Members will release their perceived or real conflict of interest at the TEP introductions on day one of the TEP. Workgroup co-chairs have been appointed for the TEP Methodology Workgroup.

Roles

Dr. Li explained that the TEP serves an advisory role to UM-KECC. They are charged with reviewing the current methodology and making recommendations on how to improve the methodology and reporting of the Star Ratings. TEP Members will be expected to share their expert opinions and experience. Their responsibility is to be a respectful team member in order for the opinions of all TEP members to be heard. Also, TEP Members are responsible for being active and engaged in the TEP discussions.

UM-KECC will serve as the moderator and facilitator of the TEP, with the goal of keeping the discussion focused and providing specific details about the Star Rating methodology and measures. UM-KECC is responsible for listening carefully and recording the TEP opinions and recommendations. They will also prepare a summary report after the TEP meeting, based on notes and audio recordings from the meeting. UM-KECC will make recommendations to CMS based on the input received from the TEP.

TEP member questions

Dr. Richard Cook: "From the discussion so far it's not clear there is any factorization of patients mixed into the evaluation, is that an accurate characterization? I guess I'm wondering when thinking about ranking different centers. Centers receiving high risk patients are going to fare poorly all other things being equal. Is the nature of the population attending a certain center factored into the evaluation of the center?"

UM-KECC responded that this question is about case mix/risk adjustment. That is regarding the measure specifications, which is not the main focus of the discussion in the TEP meeting. Risk adjustment or whether to risk adjust has been considered in the development of various measures reported on the DFC.

Dr. Jane Pendergast: “Is it correct that we are supposed to focus on the measures that have already been collected as opposed to thinking about other useful information?”

UM-KECC responded that the focus of the methodology group is on how to combine the measures into one single measure. In the joint session, the TEP members will talk about what the additional topics/elements to include in the development of the star rating.

Dr. J. Richard Landis: “Just to follow up on that, is the perspective then to evaluate the current star ranking procedure, but to also investigate how patient mix confounding could be adjusted for and potentially purpose a new ranking?”

UM-KECC responded that the focus is to talk about evaluation of the current ranking system and certainly for additional issues such as risk adjustment this is something we can talk about. The specifications of individual measures will not be the focus of the meeting. The TEP can talk about new measure topics to be included in future rating system. For more questions or information, the TEP was instructed to email UM-KECC directly. Risk adjustment/case mix has been documented in the detailed technical report for individual measures. TEP members should feel free to read them.

Dr. Catherine Sugar: “I have an administrative question. How much of an agenda structure are we going to get prior to the meeting about what the breakout and joint sessions will be? If so, when were you expecting to send the information out?”

UM-KECC responded that they will post the agenda online, so the general public will be aware of the agenda. The agenda will be sent to the TEP members very soon.

C-Tools Explanation

- UM-KECC created a web portal through the University of Michigan website, to share a variety of tools that the TEP members may find useful, such as:
 - TEP Charter and Consolidated Questions Document
 - Star Ratings Materials and full pdf versions of the articles used in the supporting literature.
- Instructions for obtaining an account were circulated to the TEP members in advance of the teleconference call, and Jordan Affholter gave a brief presentation on gaining access to the C-Tools site.
- C-Tools will be the main vehicle for providing documents to the TEP. UM-KECC encouraged TEP members to create their accounts and let UM-KECC know if any issues arise.

Public Comments

Kathy Lester: “Hi. This is Kathy Lester with Kidney Care Partners, and I want to thank Elena and Joel from CMS and all of you at UM-KECC for allowing the opportunity to have public comment. Really, I have two questions. One, I’m wondering if the background materials

you are mentioning are going to be available for interested stakeholders to access so they too can look at those prior to the in-person meeting? And then, secondly, I'm wondering how you're doing with the comment and whether any of that material that was provided prior to the establishment of the TEP is also going to be available to the members of the TEP can be provided to TEP so they can review those comments as drafted? Are you going to include the comments that were provided as well?"

UM-KECC responded that they will check with CMS to see whether that can be provided. The background materials were combined into a Consolidated Questions document, which is available to the TEP members. The Consolidated Questions document will be publicly available very soon. UM-KECC will send out the link once the document is posted. The Consolidated Questions document covers comments and responses to a number of questions from different stakeholders through the last summer and fall.

Bruce Upton: "This is Bruce Upton from Springfield, Missouri. You had a question earlier about Case Mix, but I ask if this panel is looking at different modalities and how those are different for say a PD only facility verses an in-center Hemodialysis Facility?"

UM-KECC responded that they will get back to Mr. Upton after the call with the answer to this question.

Susan Senich: "I have a question, Susan Senich from North Central PA Dialysis. It's my understanding that the current data from which the Dialysis Facility Compare and Five Star was taken from is 2013. While you are working on this methodology and everything this panel is working on-are you looking to use more current data? More current than 2013?"

UM-KECC responded that for the next release of the star ratings, which is scheduled for January 2016, would incorporate more current data; it would not be a recalculation of 2013 data.

Wrap Up

Dr. Joel Andress of CMS thanked the group for attending the teleconference. UM-KECC and CMS mentioned that the Technical Notes document addresses how to adjust for case mix on the specific measures. Dr. Li thanked the TEP members and general public for being involved in the call and requested TEP members to contact UM-KECC if they have any remaining questions.

End Stage Renal Disease Dialysis Facility Compare (DFC) Star Ratings

Technical Expert Panel
April 27 and 28, 2015



Introductions and Potential Conflict of Interest Disclosure

- UM-KECC (Facilitators)
- TEP members
- CMS/NORC



University of Michigan Kidney Epidemiology and Cost Center Star Ratings Development Team

Yi Li (Principal Investigator)

Karen Wisniewski

Joseph Messana

Christopher Harvey

Richard Hirth

Natalie Scholz

Claudia Dahlerus

Cindy Liao

Ji Zhu

Jennifer Sardone

Tempie Shearon

ZeZhi (Zac) Zhang

Casey Parrotte

Jordan Affholter



TEP Introductions

Methodology Workgroup

- Jane Pendergast (co-chair)
- Catherine Sugar (co-chair)
- Richard Cook
- J. Richard (Dick) Landis
- Franklin Maddux
- Allen Nissenon
- Dylan Small
- Nicole Stankus
- Sumi Sun



TEP Introductions

Public Reporting/Patient and Consumer Understanding Workgroup

- Paul Conway (chair)
- Joanna Galeas
- Joseph Karan
- John (Jack) Reynolds
- Tonya Saffer
- Chris Sarfaty
- David White
- “Anonymous Patient A”
– (withheld by request)



CMS/NORC Introductions

- Joel Andress, CMS
- Elena Balovlenkov, CMS
- Elaine Swift, NORC
- Rebeca Catterson, NORC



TEP Objectives

- Review the Star Rating methodology that is currently used to report Star Ratings on DFC.
- Provide input on scoring individual measures, creation of measure domains, and overall star ratings methodology.
- Make recommendations regarding measures to be added or removed, and the star ratings methodology.

TEP Member Role and Responsibilities

- The role of TEP members is to provide advisory input to UM-KECC on how to improve the methodology and reporting of star ratings
- Share your expert opinions and experience
- Be a respectful team member in order for the opinions of all TEP members to be heard
- Be engaged in discussion



UM-KECC Role and Responsibilities

- Moderate and facilitate the TEP discussion
- Listen carefully
- Dutifully and accurately record TEP opinions and recommendations in the TEP Summary Report
- UM-KECC develops advisory recommendations for CMS, based on TEP input

Agenda Day One 9:00am – 1:00pm

TEP Opening Session

| | |
|---------------|--|
| 9:00 – 9:30 | Introductions and Potential Conflict of Interest Disclosure |
| 9:30 – 9:45 | Highlight Objectives and Agenda |
| 9:45 – 10:15 | Introduction to the Methodology Presentation |
| 10:15 – 10:30 | <i>BREAK</i> |
| 10:30 – 12:00 | Workgroup Breakout Sessions |
| 12:00 – 1:00 | <i>LUNCH</i> |



Agenda Day One 1:00pm – 5:00pm

| | |
|-------------|--------------------------------|
| 1:00 – 2:20 | Continue Workgroup Discussions |
| 2:20 – 2:30 | <i>BREAK</i> |
| 2:30 – 3:45 | Continue Workgroup Discussions |
| 3:45 – 4:00 | <i>BREAK</i> |
| 4:00 – 5:00 | Joint TEP Discussion |



Agenda Day 2 - 9:00am – 3:00pm

| | |
|---------------|--------------------------------|
| 9:00 – 10:15 | Continue Workgroup Discussions |
| 10:15 – 10:30 | <i>BREAK</i> |
| 10:30 – 12:00 | Continue Workgroup Discussions |
| 12:00 – 1:00 | <i>LUNCH</i> |
| 1:00 – 2:30 | TEP Concluding Session |
| 2:30 – 3:00 | Public Comment Period |



Overview of Methodology

Background: Medicare.gov Compare Websites

- Official CMS source for information on health care provider quality
- Quality measures based on established scientific standards of rigor and accuracy
- Five public reporting sites:
 - Nursing Home Compare (1999)
 - ***Dialysis Facility Compare*** (2001)
 - Home Health Compare (2005)
 - Hospital Compare (2005)
 - Physician Compare (2010)



Expansion of Dialysis Facility Compare Reported Quality Measures

| 2001 - 2012 | 2013 | 2014 | 2015 |
|----------------|---------------------------|---------------------------|---------------------------|
| SMR | SMR | SMR | SMR |
| URR | SHR | SHR | SHR |
| Hgb >12.0 g/dL | URR | STrR | STrR |
| Hgb <10.0 g/dL | <i>Hgb >12.0 g/dL</i> | URR | SRR** |
| | <i>Hgb <10.0 g/dL*</i> | <i>Hgb >12.0 g/dL</i> | <i>Hgb >12.0 g/dL*</i> |
| | Fistula | <i>Hgb <10.0 g/dL*</i> | <i>Hgb <10.0 g/dL*</i> |
| | Catheter | Fistula | Fistula |
| | Adult HD Kt/V | Catheter | Catheter |
| | Adult PD Kt/V | Adult HD Kt/V | Adult HD Kt/V |
| | Pediatric HD Kt/V | Adult PD Kt/V | Adult PD Kt/V |
| | | Pediatric HD Kt/V | Pediatric HD Kt/V |
| | | Hypercalcemia | Hypercalcemia |
| | | <i>Serum Phosphorus*</i> | <i>Serum Phosphorus*</i> |

*These measures were included in the DFC reports released to facilities and the DDB files but were not displayed on the DFC site

**SRR was included in preview reports and appeared on the 2015 April website refresh. Not included in star rating.

URR, Hgb, and Serum Phosphorus not included in the calculations for the January 2015 Star Ratings release.



Why Star Ratings?

- Mandate for increased transparency, wider and easier use of publicly reported data on health care quality
 - Medicare Improvements for Patients and Providers Act (2008)
 - Affordable Care Act (2010)
 - National Quality Strategy (2011)
 - Obama Administration's Digital Government Strategy (2012)



Star Ratings on Compare Websites

- Response to calls for increased sharing of information and more effective public reporting
- Current Star Rating programs:
 - Nursing Home Compare
 - Medicare Plan Finder
 - Physician Compare
 - Dialysis Facility Compare
 - Hospital Compare
- Upcoming:
 - Home Health Compare



DFC Star Ratings and Consumers

- Provides an easily recognizable and understandable way to compare facilities
- Offers additional information that consumers can use to make better informed decisions or identify issues, along with:
 - Visiting the facility and asking questions
 - Talking with a doctor
 - Looking up data on individual quality measures



DFC Star Ratings: What Do They Mean?



Much Above Average



Above Average



Average



Below Average



Much Below Average



Star Rating Methodology

DFC Quality Measures Used in Rating

Risk adjusted outcome measures:

- Standardized Mortality Ratio (SMR): lower is better
- Standardized Hospitalization Ratio (SHR): lower is better
- Standardized Transfusion Ratio (STrR): lower is better



DFC Quality Measures Used

Intermediate outcome measures:

- **Hypercalcemia:** Percentage of adult patients who had high calcium in blood over the past 3 months (lower is better)
- **Fistula:** Percentage of adult patients who received dialysis through arteriovenous fistula (higher is better)
- **Catheter:** Percentage of adult patients who had a catheter left in a vein \geq 90 days for regular hemodialysis (lower is better)



DFC Quality Measures Used

Intermediate outcome measures (continued):

- **Dialysis Adequacy:** Percentage of patients who had enough waste removed from blood during dialysis
 - Dialysis Adequacy is measured by Kt/V
 - For the Star Ratings, the adult HD, adult PD, and pediatric HD dialysis adequacy measures are combined and scored as one measure:
 - **Adult hemodialysis (HD) Adequacy** (higher is better)
 - **Pediatric hemodialysis (HD) Adequacy** (higher is better)
 - **Adult peritoneal dialysis (PD) Adequacy** (higher is better)



DFC Quality Measures

- All measures have been developed based on input from previous clinical TEPs
- Risk adjustment or whether to risk adjust for individual measures has gone through rigorous TEP processes in the past 10 years

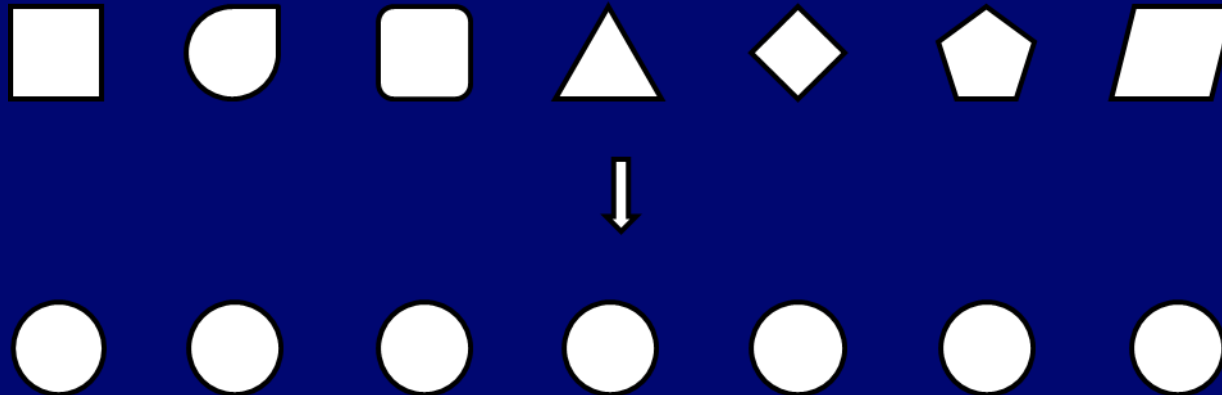


Summary of DFC Star Ratings



**7 Measures with different scales, averages,
distributions, direction**

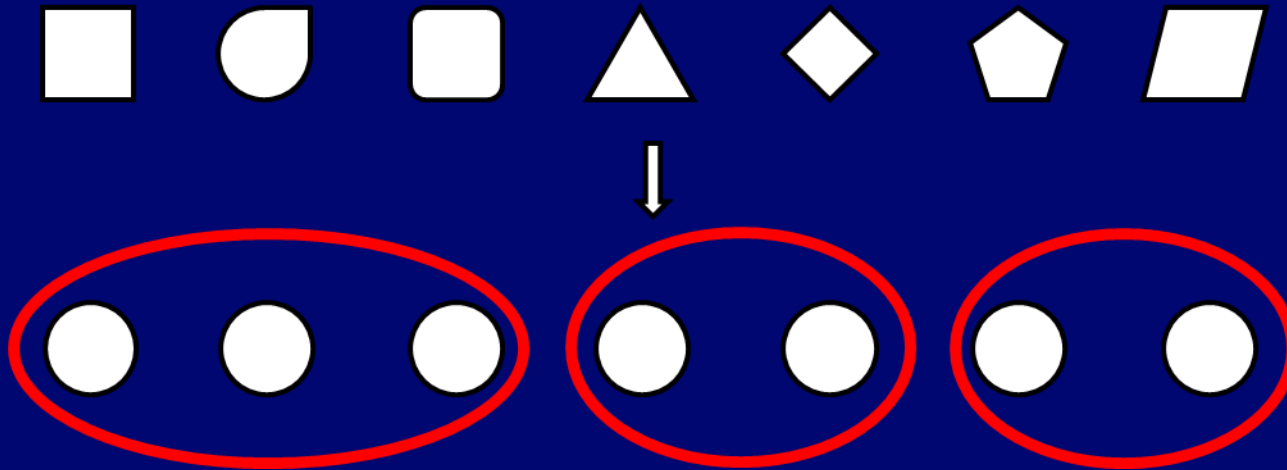
Summary of DFC Star Ratings



Score measures to make them comparable

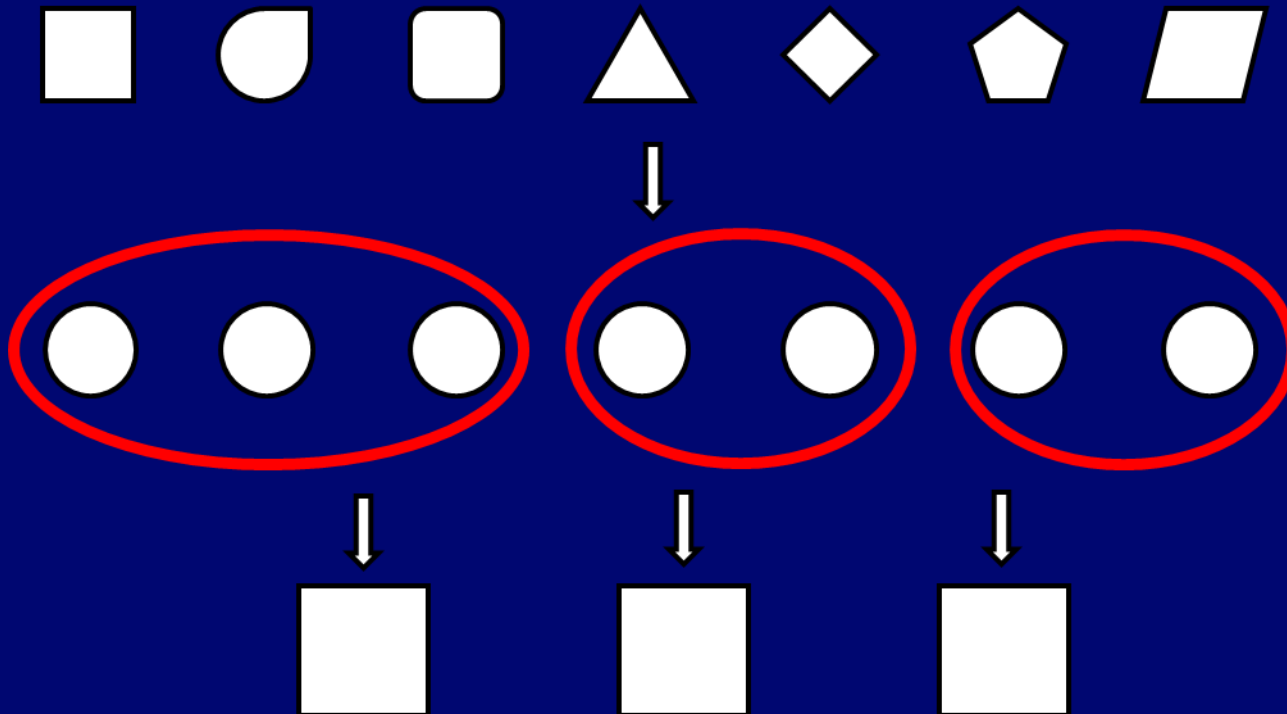


Summary of DFC Star Ratings



**Score measures grouped based on
correlation structure
(most related measures in same group)**

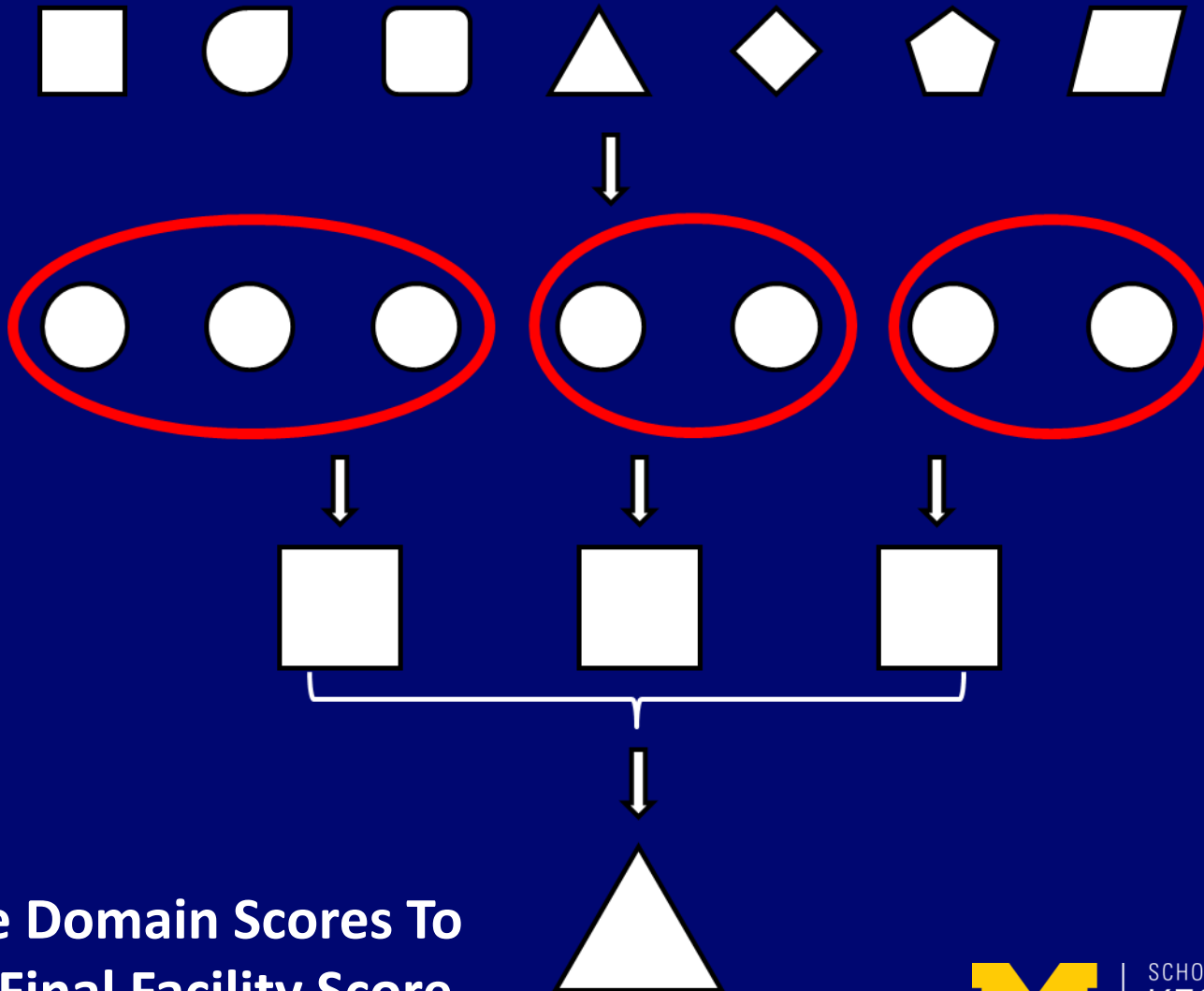
Summary of Ratings



**Average score measures within groups to create
Domain Scores**



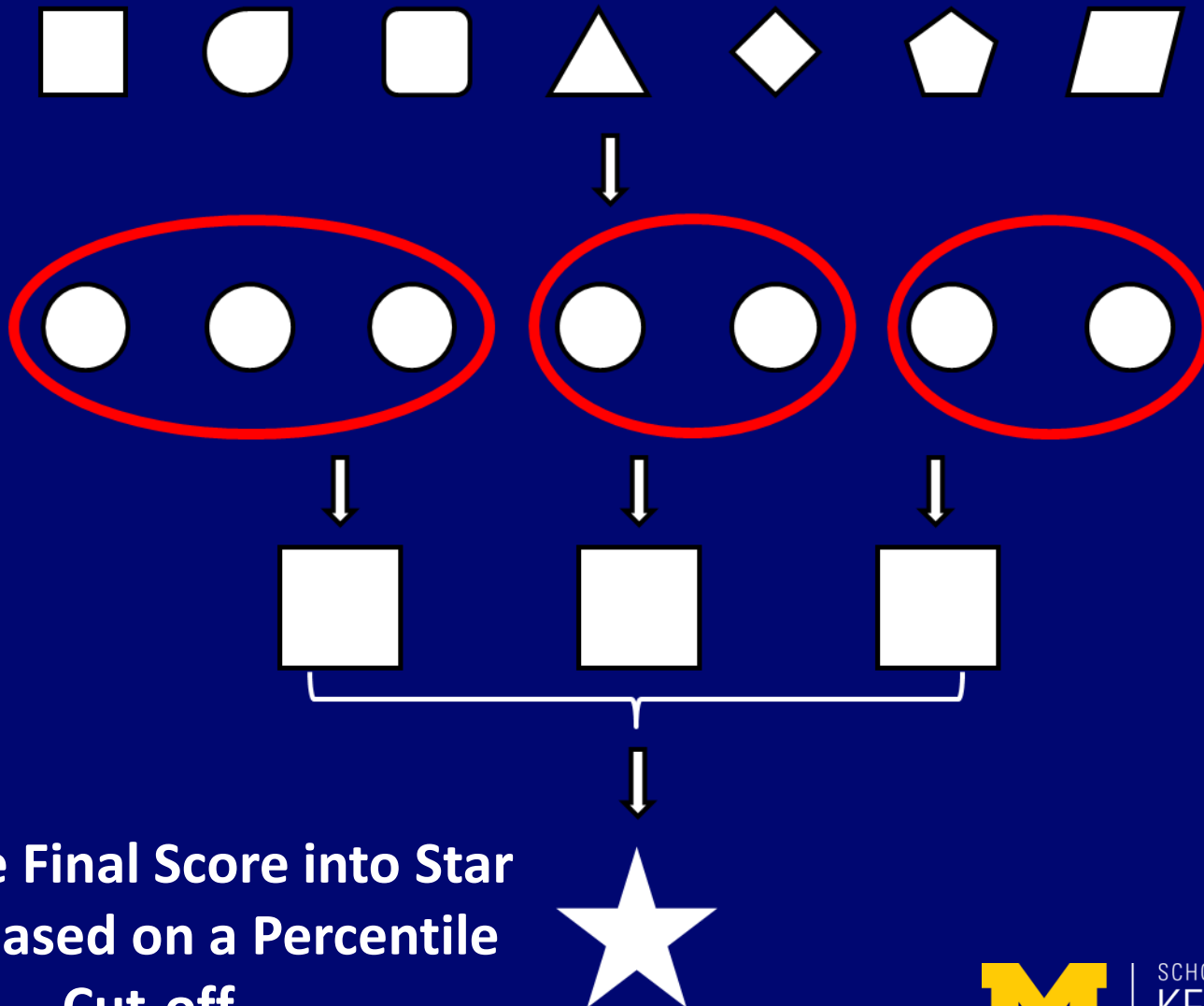
Summary of Ratings



**Average Domain Scores To
Create Final Facility Score**



Summary of Ratings



**Translate Final Score into Star
Rating Based on a Percentile
Cut-off**



Example of Star Rating

| | | | | | | | |
|----------------|------------|------------|-------------|----------------|-----------------|-------------|-----------------|
| Measure | SMR | SHR | STrR | Fistula | Catheter | Kt/V | Hypercal |
| Values: | 1.4 | 1.1 | 1.6 | 58% | 10% | 93% | 2% |



Example of Star Rating

| Measure | SMR | SHR | STrR | Fistula | Catheter | Kt/V | Hypercal |
|---------|-----|-----|------|---------|----------|------|----------|
| Values: | 1.4 | 1.1 | 1.6 | 58% | 10% | 93% | 2% |

Star Rating: 



Example of Star Rating

| Measure | SMR | SHR | STrR | Fistula | Catheter | Kt/V | Hypercal |
|---------|-----|-----|------|---------|----------|------|----------|
| Values: | 1.4 | 1.1 | 1.6 | 58% | 10% | 93% | 2% |

Ranks:
(0-100)

23

40

27

42

48

49

41

Domain Scores:
(0-100)

30

44

50

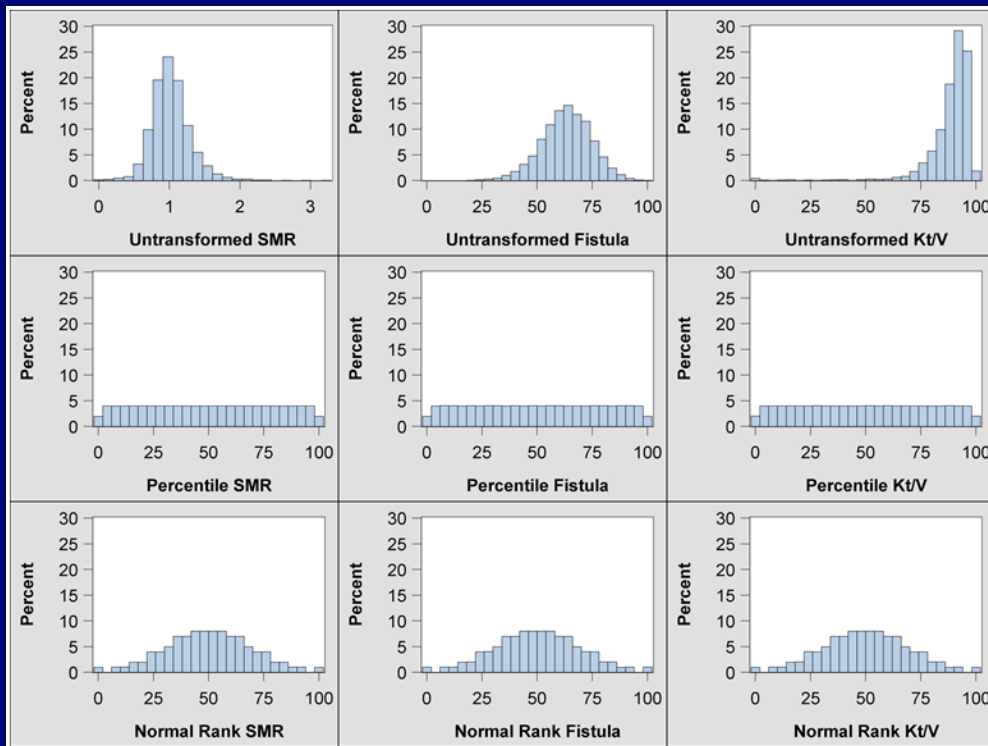
Final Score:
(0-100)

42

Star Rating:



Measure Scoring



- Rank measures to make them have the same scales
- Probit transform measures to make them comparable



Determination of Measure Domains

- Domains based on grouping of measures with highest correlations (Spearman)
- Grouped with the aid of factor analysis

| Measures | STrR | SHR | SMR | Adequacy | Hypercal | Fistula | Catheter |
|----------|------|-------------|-------------|----------|-------------|---------|-------------|
| STrR | 1.00 | 0.40 | 0.21 | 0.08 | 0.00 | 0.11 | 0.15 |
| SHR | | 1.00 | 0.26 | 0.11 | 0.01 | 0.13 | 0.19 |
| SMR | | | 1.00 | 0.08 | 0.05 | 0.17 | 0.11 |
| Adequacy | | | | 1.00 | 0.19 | 0.06 | 0.13 |
| Hypercal | | | | | 1.00 | 0.09 | 0.05 |
| Fistula | | | | | | 1.00 | 0.45 |
| Catheter | | | | | | | 1.00 |



Measure Domains

- “*Standardized Outcomes (SMR, SHR, STrR)*” which includes measures for mortality, hospitalization, and transfusions
- “*Other Outcomes 1 (AV fistula, tunneled catheter)*” which includes the arteriovenous fistula and catheter > 90 days measures
- “*Other Outcomes 2 (Kt/V, hypercalcemia)*” which includes the combined HD and PD adequacy measure (Kt/V) and the hypercalcemia measure

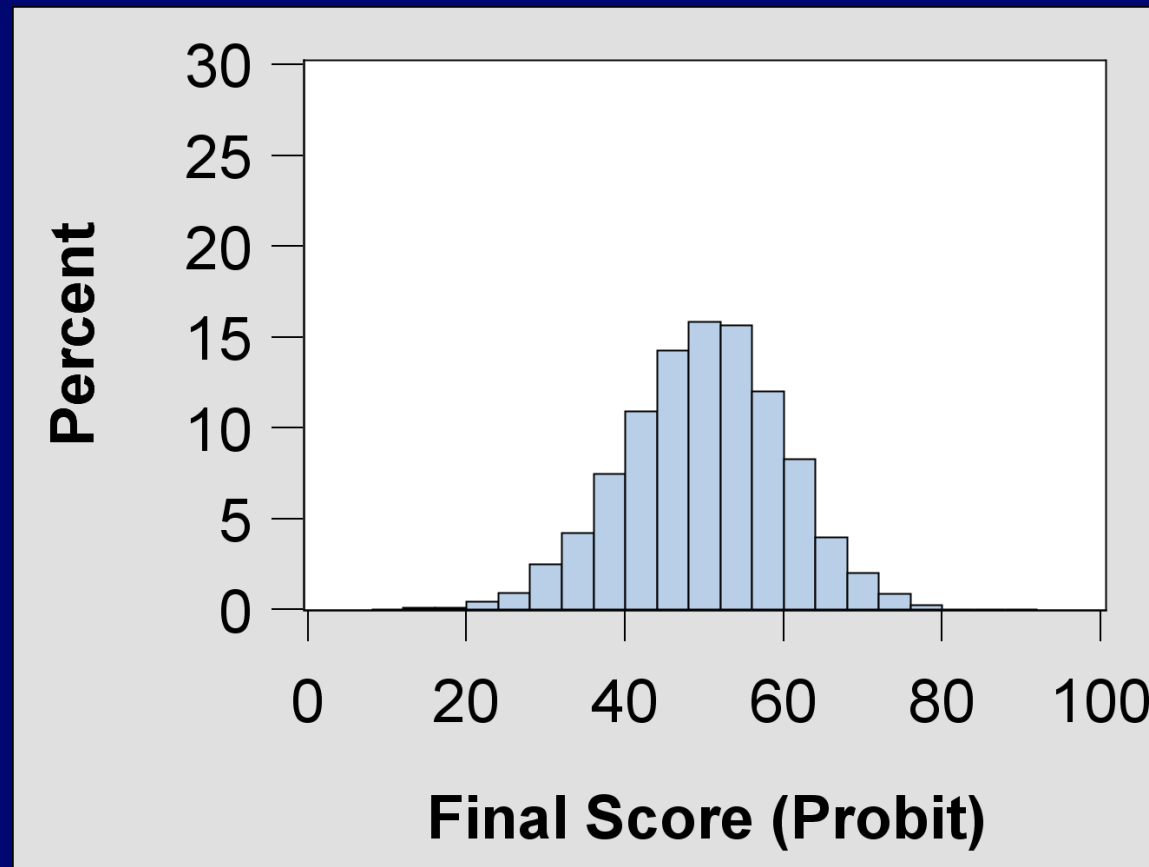


Missing Measure Values

- Facilities missing an entire domain are not given a Star Rating
- Exception: PD-Only dialysis facilities will be scored and given a star rating by averaging only the “Standardized Outcomes” and “Other Outcomes 2” domains
 - This is because PD – Only facilities do not report measures for the Fistula/Catheter domain and are therefore only rated based on the other 2 domains
- For all other facilities, if no domains are missing, the missing measure(s) within a domain are imputed with the national average rank for that measure (probit rank of 50)



Distribution of Final Scores



Star Rating Cut-Offs

- Facilities ranked highest to lowest based on final scores, then binned into star rating categories

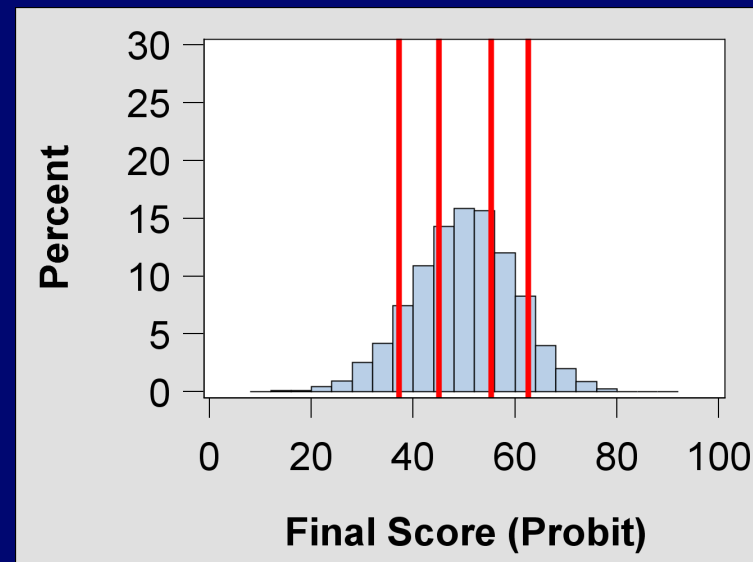
10% 5 stars

20% 4 stars

40% 3 stars

20% 2 stars

10% 1 star



Summary

- Methodologies developed to combine individual measures with different scales and distributions
- Factor analysis determined domains based on statistical association of measures
- Average of domain scores determined overall score for star rating category
- Star rating provides summarized information in addition to individual measures



Public Reporting/Patient and Consumer Understanding Presentation

Workgroup Morning Session

- Description of DFC Star Rating measures (SMR, SHR, STrR, vascular access, Kt/V, hypercalcemia)
- Continued discussion from the General Methodology Presentation
 - Questions or clarifications?
- Discussion



Mortality and Eye Color- Hypothetical Model for Mortality

- Researchers conducted a comprehensive study of mortality associated with facelifts in ambulatory surgery facilities
- Overall, peri-operative mortality was 7.5% (national average)
- Based on national data, “blue eyes” had 10% mortality; “all other eyes” had 5% mortality
- National data- 50% of participants had blue eyes and 50% had “other” eye color.

Mortality and Eye Color- Hypothetical Model for Mortality

| | Risk Distribution | Mortality | Risk Adj. Mort (%) | SMR |
|------------|---------------------|-----------|---------------------------|---------------|
| Facility A | 100% blue eyes | 20% | 15% ($20/10 * 7.5$) | 2.0 (20/10) |
| Facility B | 100% other eyes | 10% | 15% ($10/5 * 7.5$) | 2.0 (10/5) |
| Facility C | 100% blue eyes | 7.5% | 5.625% ($7.5/10 * 7.5$) | 0.75 (7.5/10) |
| Facility D | 50% blue/ 50% other | 10% | 13.3% ($10/7.5 * 7.5$) | 1.33 (10/7.5) |

DFC Standardized Measures in Star Rating

- Standardized Mortality Ratio (SMR): patient death rate (lower is better)
- Standardized Hospitalization Ratio (SHR): rate of hospital admissions (lower is better)
- Standardized Transfusion Ratio (STrR): rate of transfusions (lower is better)
- Risk adjusted for patient characteristics
- SMR, SHR, and STrR measures based on previous TEPs (Technical Expert Panels)



Standardized Mortality Ratio (SMR)

- The SMR is the actual death rate divided by the expected death rate
- Actual death rate: rate at which patients in a facility died during the period indicated on the table
- Expected death rate: rate at which patients are expected to die given the facility patient case-mix
- On DFC, a facility's mortality rate is rated as Better than Expected, As Expected, or Worse than Expected



Risk Adjustment of the Standardized Mortality Ratio (SMR)

- SMR is calculated for all dialysis patients, both Medicare and non-Medicare
- Expected death rate takes into account certain patient characteristics: patient age, race, ethnicity, sex, diabetes, duration of ESRD, nursing home status, patient comorbidities at ESRD incidence, body mass index (BMI) at ESRD incidence, and population death rates



Standardized Hospitalization Ratio (SHR)

- The SHR is the actual hospitalization rate divided by the expected hospitalization rate
- Actual hospitalization rate: rate at which patients in a facility were hospitalized during the period indicated on the table
- Expected hospitalization rate: rate at which patients are expected to be hospitalized given the facility patient case-mix
- On DFC, a facility's hospital admission rate is rated as Better than Expected, As Expected, or Worse than Expected



Risk Adjustment of the Standardized Hospitalization Ratio (SHR)

- SHR includes only Medicare patients
- Expected hospitalization rate takes into account certain patient characteristics: patient age, sex, diabetes, duration of ESRD, nursing home status, patient comorbidities at ESRD incidence, BMI at ESRD incidence, and calendar year



Standardized Transfusion Ratio (STrR)

- The STrR measure is the ratio of observed red blood cell transfusion events occurring in patients dialyzing at a facility, to the expected rate under a national norm
- Expected number of transfusion events
 - Number of eligible red blood cell transfusion events that would be expected among patients at a facility during the reporting period, given the facility's patient mix
- Observed number of transfusion events
 - Number of actual eligible red blood cell transfusion events occurring in patients who dialyze at the facility during the reporting period
 - Patients are excluded from the calculation if certain co-morbidities have been identified from Medicare claims within the prior year.

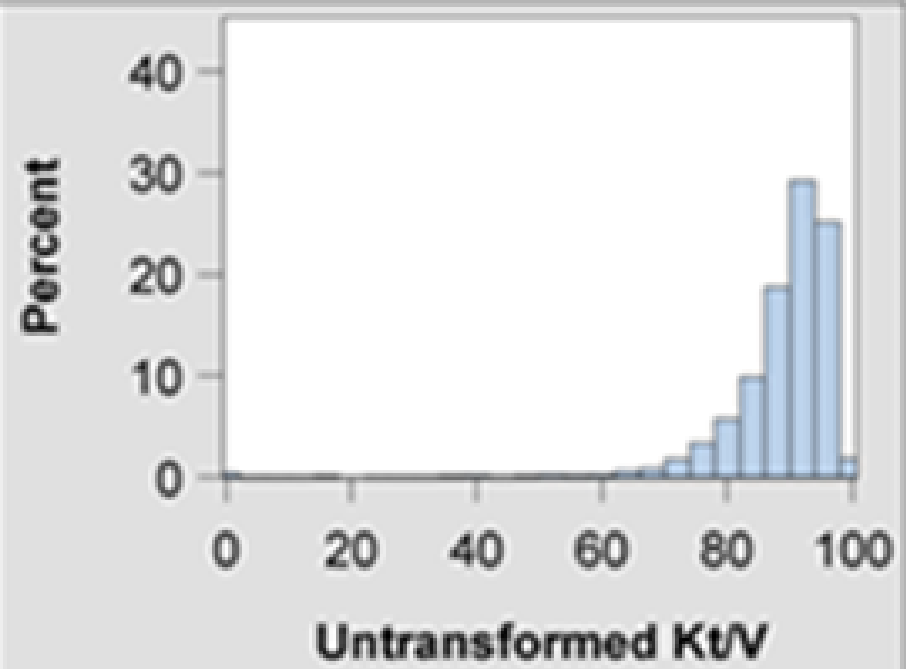
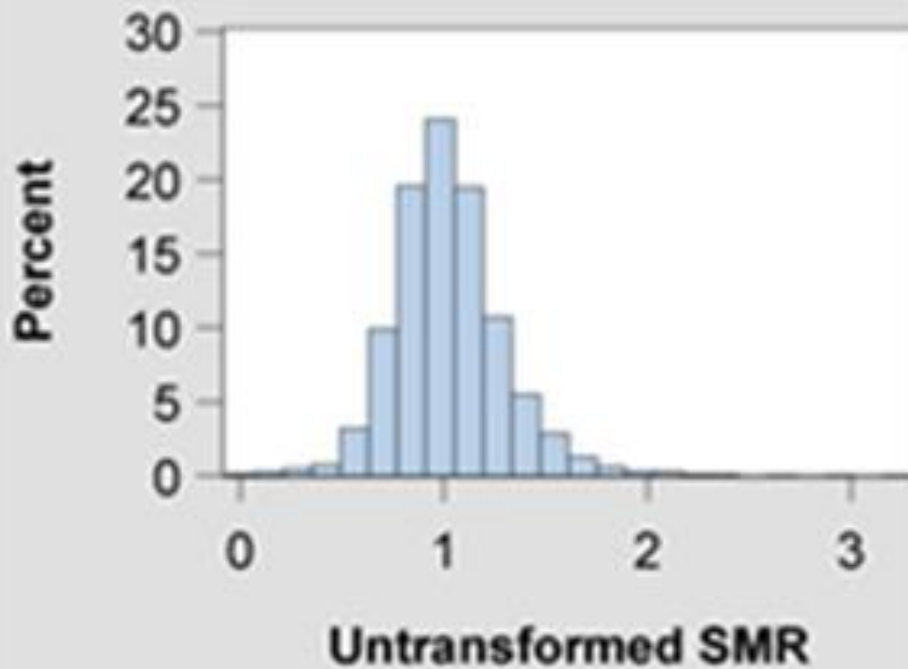


Risk Adjustment of the Standardized Transfusion Ratio (STrR)

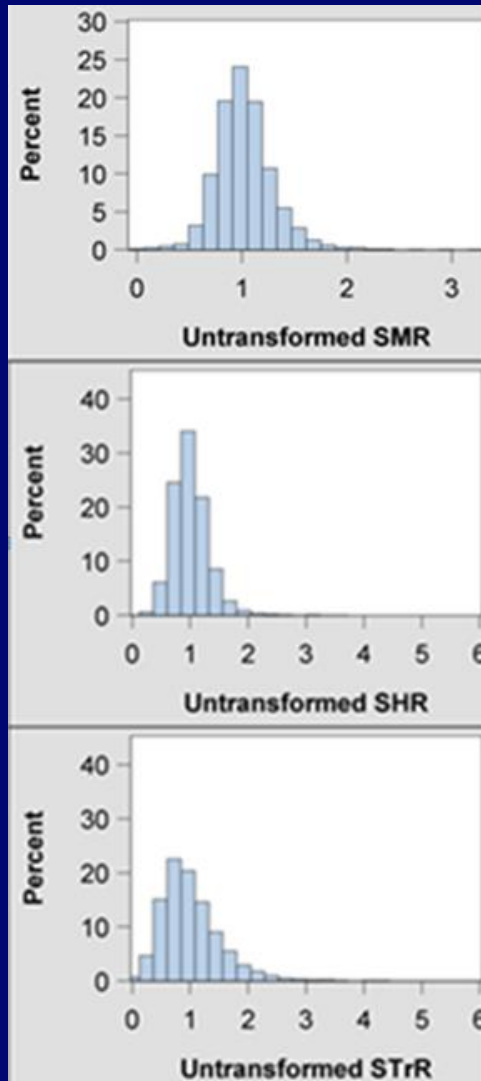
- STrR includes only Medicare patients
- Expected transfusion event rate takes into account certain patient characteristics: patient age, sex, diabetes, duration of End-Stage Renal Disease (ESRD), nursing home status, patient comorbidities at ESRD incidence, BMI at ESRD incidence, and calendar year.



Different Measure Distributions and Numerical Scales



Quality Measure Distributions

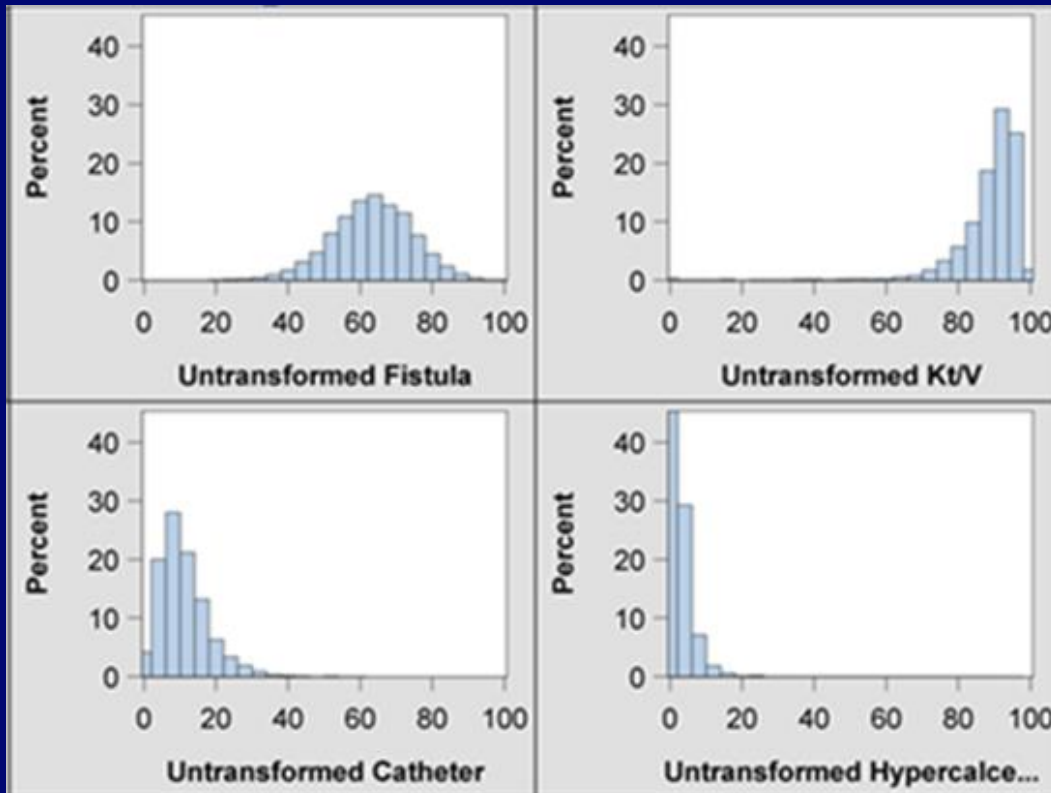


Distribution of the Untransformed Standardized Measures

- Standardized Mortality Ratio (SMR)
- Standardized Hospitalization Ratio (SHR)
- Standardized Transfusion Ratio (STrR)



Quality Measure Distributions

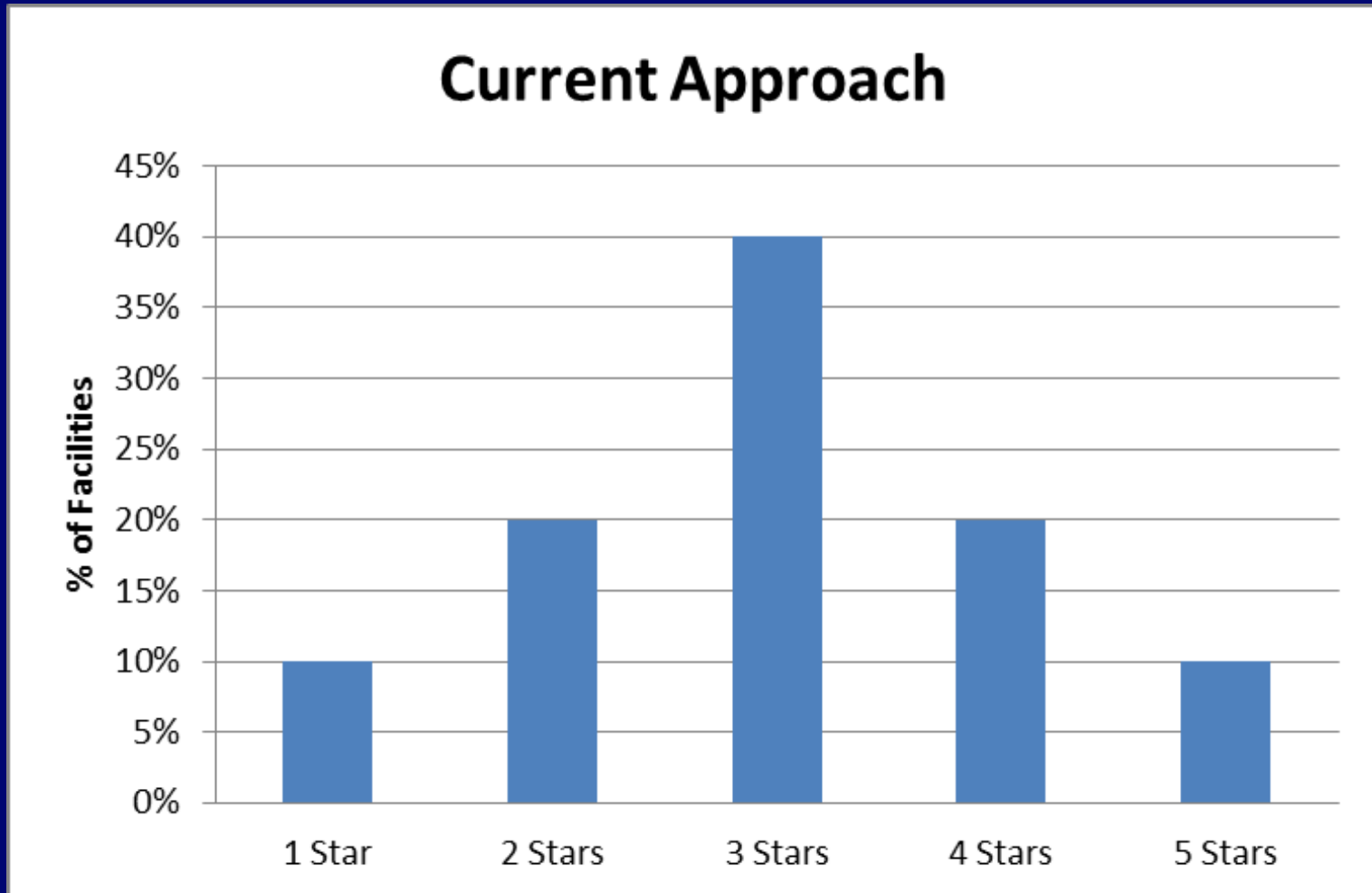


Distribution of Untransformed Intermediate Outcome Measures

- Fistula
- Catheter
- Kt/V
- Hypercalcemia



Current Star Rating Distribution



Comparisons of Averages of Individual Measures between DFC Star Rating Categories

| Measure | 1 Star | 2 Star | 3 Star | 4 Star | 5 Star |
|--------------------|--------|--------|--------|--------|--------|
| STrR | 1.52 | 1.19 | 0.98 | 0.82 | 0.62 |
| SMR | 1.34 | 1.11 | 1.01 | 0.93 | 0.83 |
| SHR | 1.28 | 1.12 | 0.99 | 0.87 | 0.76 |
| All Kt/V | 79.10 | 85.24 | 89.26 | 92.38 | 94.45 |
| Hypercalcemia | 4.63 | 3.34 | 2.28 | 1.24 | 0.83 |
| AVF | 50.24 | 57.95 | 63.35 | 68.97 | 74.69 |
| Catheter > 90 days | 20.21 | 14.26 | 10.22 | 7.27 | 5.12 |

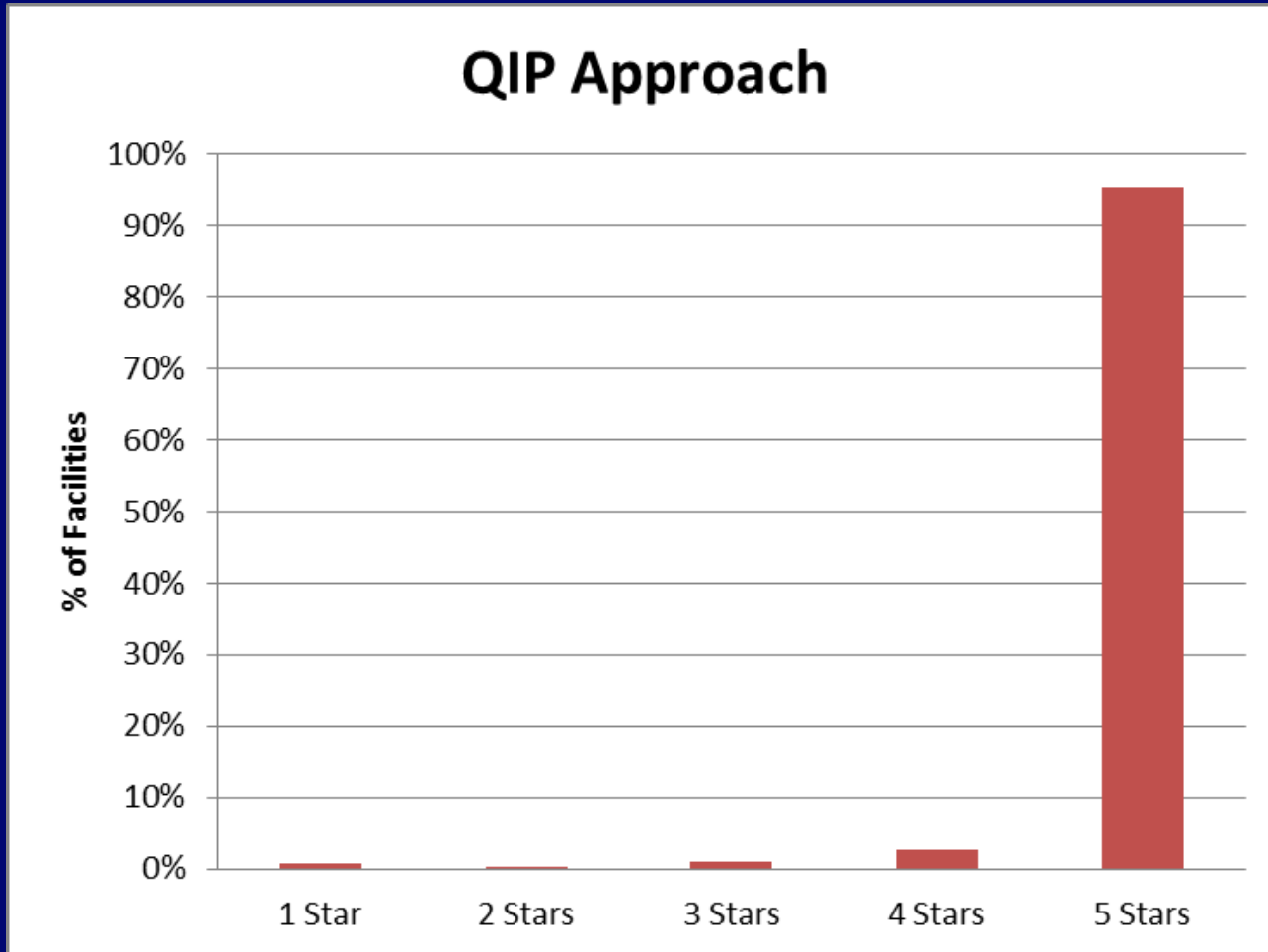
-For STrR, SMR, SHR, hypercalcemia, and catheter, lower values are better; for all Kt/V and AVF, higher values are better.

-There are highly statistically significant and clinically meaningful differences between these mean values in adjacent tiers

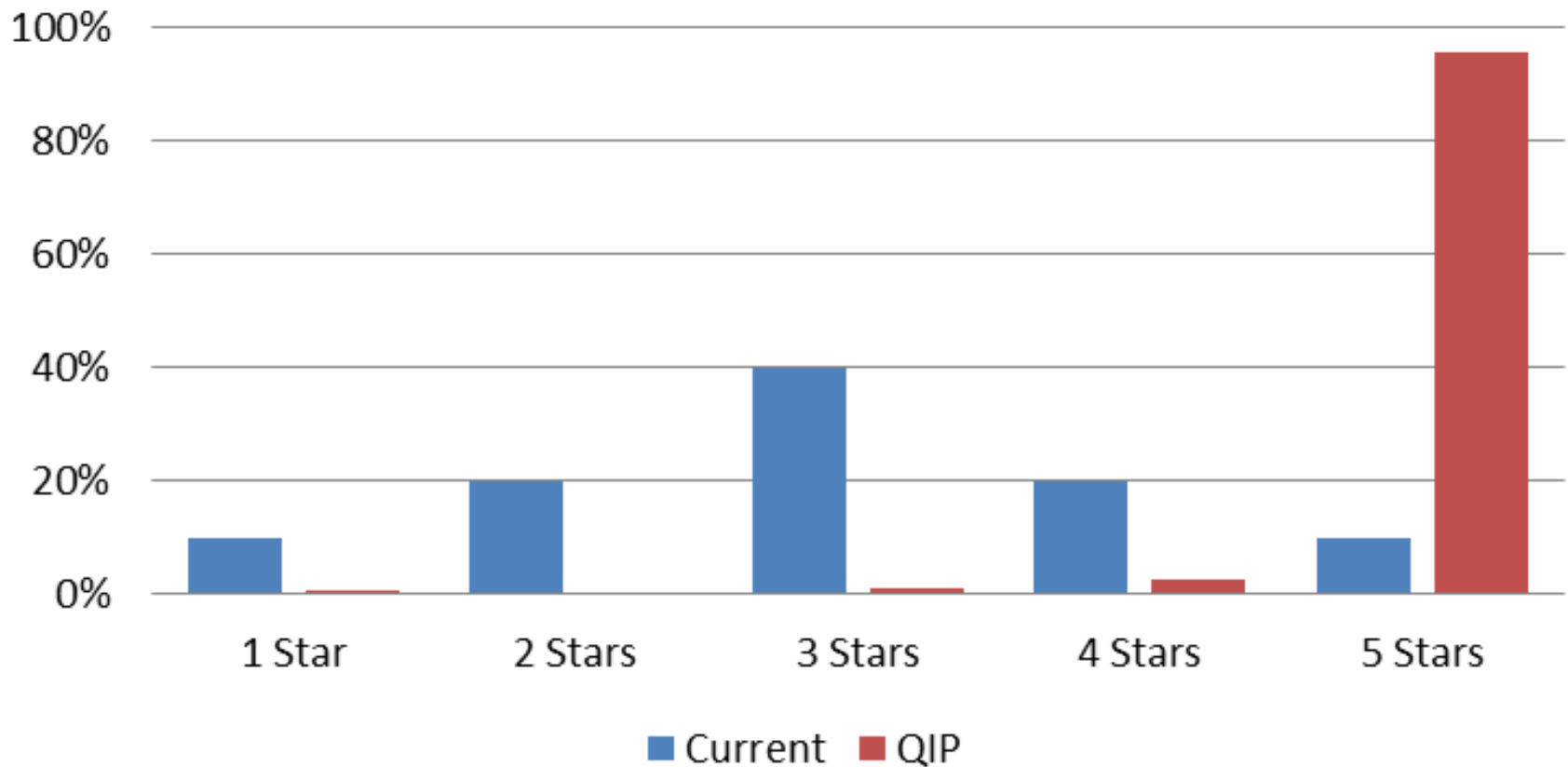
-Table is from the Consolidated Questions Document (Table 2)



Star Ratings Based on QIP Payment Reduction Categories

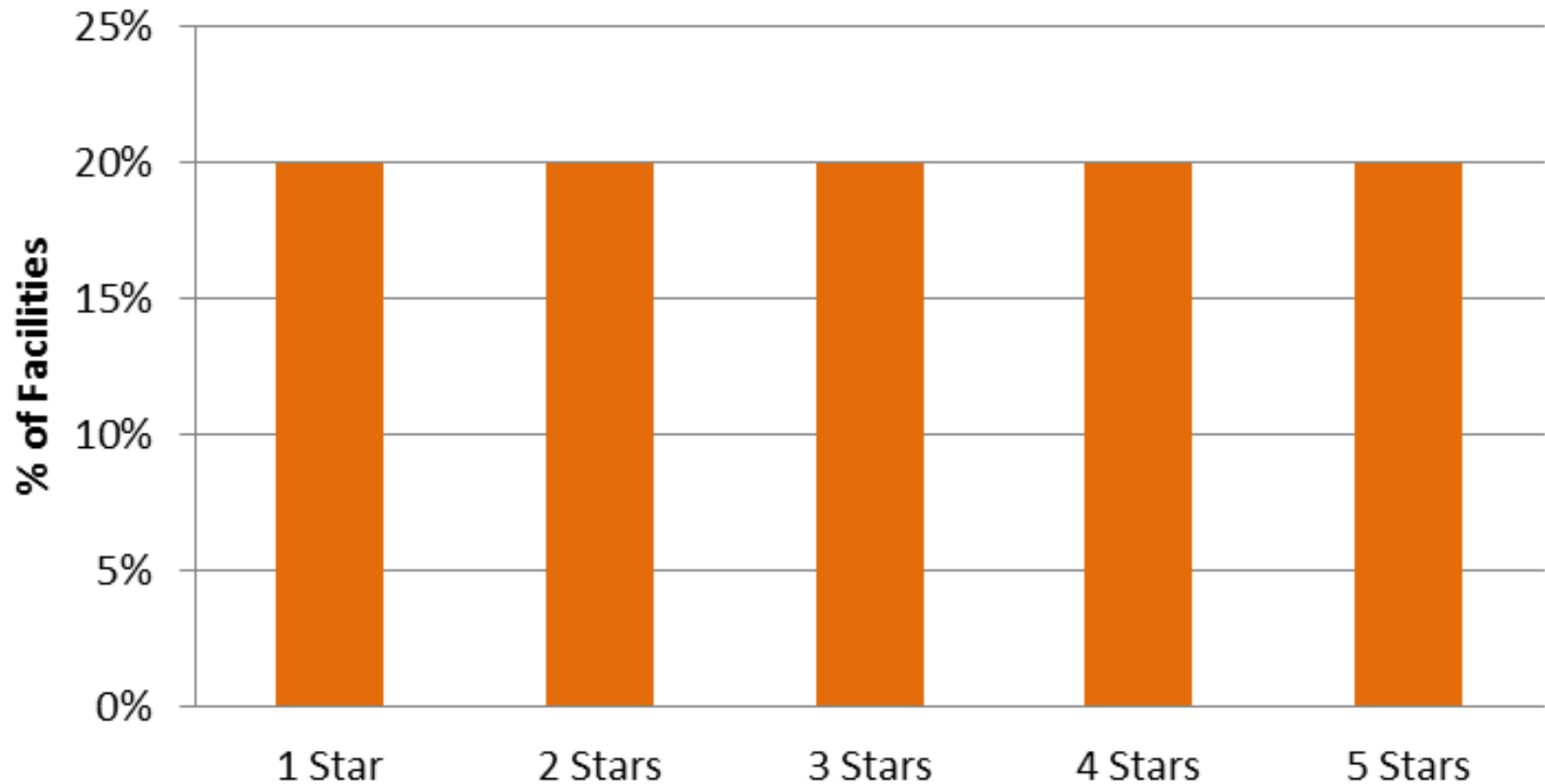


Current Approach vs. QIP

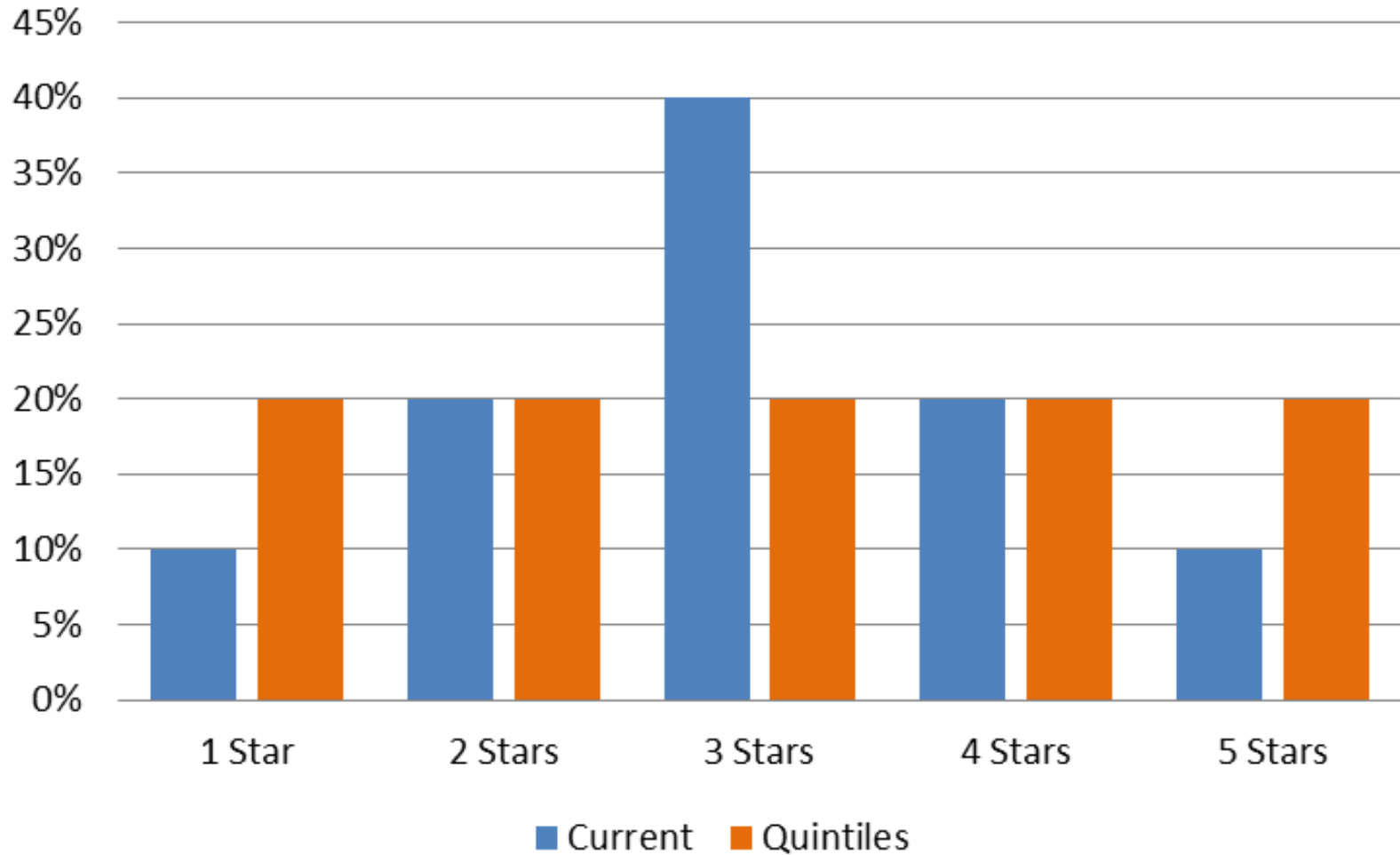


Star Ratings Based on Equal Quintiles

Equal Quintiles Approach



Current Approach vs. Quintiles



Discussion



Absolute vs. Relative Rankings

- Dr. Messina is scheduled to review the applications and interview four medical students from two different medical schools (Med School “A” and Med School “B”). Based on these activities, he must choose two of the students for admission to his medical residency training program.
- He wants to choose the best candidates because much depends on upon it (his happiness, the health and satisfaction of his patients, the reputation of his medical center, etc.)
- Med School A grades students on an absolute scale. The transcripts from this school reveal that 85% of students are graded as “Honors”, the highest grade possible. Most of the remaining students receive “High Pass” and only a rare student “fails”
- Med School B grades students on a strict curve. Each year 10% of students graduate with “Honors”, 20% with “High Pass”, 40% with “Pass”, 20% with “Marginal Pass” and 10% “Fail”



Absolute vs. Relative Rankings

| Student | #1 | #2 | #3 | #4 |
|------------|------------|------------|------------|------------|
| Med School | School "A" | School "A" | School "B" | School "B" |
| Grade | Honors | Honors | Honors | High Pass |
| Interview | Good | Good | Good | Good |
| Final Rank | ??? | ??? | ??? | ??? |



Consistency in Ranking Across Public Reporting Programs

- Should there be one method for reporting quality for all programs
 - For example, same way of calculating overall score for QIP and DFC 5 star?



Potential Unintended Consequences of the Current Star Rating Program

- Involuntary discharge of patients
- “Cherry picking” patients
- Potentially creating false differences between facilities
- Impact of geographic variation/disparities
- Patient concerns with local facility rating choices
- Other potential consequences



Public Reporting/Patient and Consumer Understanding Day Two



Day 2 Discussion

- Review of Day One Discussions
- Patient and consumer education for star ratings

Public Reporting/Patient and Consumer Understanding




Public Reporting/Patient and Consumer Understanding

- What are the best methods for educating the public about the DFC Star Ratings?
 - Examples: Patient Centered Open Door Forums/Presentations, Written materials, Videos, Other
- What level of content detail is useful in these outreach and education methods?



What are the Star Ratings



The Official U.S. Government Site for Medicare

Dialysis Facility Compare

About Dialysis Facility Compare

What is Dialysis Facility Compare?

» What information can I get about dialysis facilities?

» **Star ratings**

Dialysis facility characteristics and services

Quality measures

How do I use the information on Dialysis Facility Compare?

About the data

Resources

Help

[Back to Top](#)

What are the star ratings

What are the star ratings?

The star ratings are part of Medicare's efforts to make data on dialysis centers easier to understand and use. The star ratings show whether your dialysis center provides quality dialysis care – that is, care known to get the best results for most dialysis patients. The rating ranges from 1 to 5 stars. A facility with a 5-star rating has quality of care that is considered 'much above average' compared to other dialysis facilities. A 1- or 2- star rating does not mean that you will receive poor care from a facility. It only indicates that measured outcomes were below average compared to those for other facilities.

Where do the star ratings come from?

The star ratings use several measures reported on Dialysis Facility Compare. These measures reflect the quality of care at each dialysis center. [Get more detail about where the star ratings come from.](#)

How can I use the star ratings?

If you are new to dialysis, you and your doctor can talk about what the ratings mean and how you can use them with other information to help you decide where to go for treatment. We also suggest you visit a dialysis unit, look around, meet the staff, and [ask questions](#).



Star Ratings Overview

Medicare.gov
Dialysis Facility Compare

The Official U.S. Government Site for Medicare

About Dialysis Facility Compare

About the data

About the Data

What measures are displayed?

» Star ratings

Quality measures: best treatment practices

Quality measures: hospitalizations and deaths

Footnotes

Data sources

Data collection periods and updates

Download the data

Resources

Help

Star ratings

Overview

The [Dialysis Facility Compare \(DFC\)](#) website provides each facility with a star rating. This gives patients with End-stage Renal Disease (permanent kidney failure requiring dialysis or a kidney transplant), their families, and caregivers more information on quality of care. It also makes it easier to compare facilities.

The [Centers for Medicare and Medicaid Services \(CMS\)](#) and the [University of Michigan Kidney Epidemiology and Cost Center](#) have developed a methodology for rating each facility. The method produces a final score that is based on quality measures currently reported on the DFC website.

The rating ranges from 1 to 5 stars. A facility with a 5-star rating has quality of care that is considered 'much above average' compared to other dialysis facilities. A 1- or 2- star rating does not mean that you will receive poor care from a facility. It only indicates that measured outcomes were below average compared to those for other facilities. Star ratings on Dialysis Facility Compare are updated annually to align with the annual updates of the standardized measures.

The quality measures used in the star ratings

The Dialysis Facility Compare website currently reports on 9 measures of quality of care for facilities. These measures are used to develop the star rating:

1. Standardized Hospitalization Ratio (SHR)
2. Standardized Mortality Ratio (SMR)
3. Standardized Transfusion Ratio (STrR)
4. Percentage of adult hemodialysis (HD) patients who had enough wastes removed from their blood during dialysis: Kt/V greater than or equal to 1.2
5. Percentage of pediatric hemodialysis (HD) patients who had enough wastes removed from their blood during dialysis: Kt/V greater than or equal to 1.2
6. Percentage of adult peritoneal dialysis (PD) patients who had enough wastes removed from their blood during dialysis: Kt/V greater than or equal to 1.7
7. Percentage of adult dialysis patients who had an average calcium over the past three months greater than 10.2 mg/d (hypercalcemia)
8. Percentage of adult patients who received treatment through arteriovenous fistula (AVF)
9. Percentage of adult patients who had a catheter (tube) left in a vein longer than 90 days, for their regular hemodialysis treatment (catheter > 90)



Star Ratings Description

Quality measure domains

The measures used in the star rating are grouped into three domains by using a statistical method known as Factor Analysis. Each domain contains measures that are most correlated. This allows CMS to weight the domains rather than individual measures in the final score, limiting the possibility of overweighting quality measures that assess similar qualities of facility care. The three domains are as follows:

- ◆ **"Standardized Outcomes (SHR, SMR, and STrR)"** – This first domain combines the three outcome measures for hospitalization, mortality and transfusions (SHR, SMR, and STrR).
- ◆ **"Other Outcomes 1 (AV fistula, tunneled catheter)"** – The arteriovenous fistula and catheter measures forms the second domain.
- ◆ **"Other Outcomes 2 (Kt/V, hypercalcemia)"** – The All Kt/V and hypercalcemia measures forms the third domain.

Facilities are rated as long as they have at least one measure in each of the three domains. Because the vascular access measures in the "Other Outcomes 1 (AV fistula, tunneled catheter)" domain do not apply to peritoneal dialysis patients, peritoneal dialysis-only facilities are rated based on the other two domains. They receive ratings as long as they have scores for at least one of the two domains not related to vascular access.

The star rating

To calculate the star rating for a facility, each domain is given a domain score between 0 and 100 by averaging the normalized scores for measures within that domain. A final score between 0 and 100 is obtained by averaging the three domain scores (or two domain scores for peritoneal dialysis-only facilities). Finally, to recognize high and low performances, facilities receive stars in the following way:

- ◆ Facilities with the top 10% final scores were given a star rating of 5.
- ◆ Facilities with the next 20% highest final scores were given a star rating of 4.
- ◆ Facilities within the middle 40% of final scores were given a star rating of 3.
- ◆ Facilities with the next 20% lowest final scores were given a star rating of 2.
- ◆ Facilities with the bottom 10% final scores were given a star rating of 1.

The results

The star ratings detect the differences in facility performance based on the measures reported on the DFC website. In the scoring algorithm using the January 2014 data, the averages of individual measures are consistently better with higher star ratings.

Further information about the content and scoring of each domain, as well as the star rating can be found in the [Technical Users' Guide](#).



Star Ratings on Dialysis Facility Compare

Compare dialysis facilities

[Back to Results](#)


General information

Best treatment practices

Hospitalizations & deaths

UNIV OF MI DIALYSIS - ADULT x

1500 E MEDICAL CTR DR
D7605 UNIVERSITY HOSPITAL
ANN ARBOR, MI 48109
(734) 936-5656

Rating : ★●●●●

Distance : 4.3 miles

[Add to my Favorites](#)
[Map and Directions](#)

MICHIGAN DIALYSIS SERVICES OF ANN ARBOR x

2850 S INDUSTRIAL HWY STE 100
ANN ARBOR, MI 48104
(734) 677-1490

Rating : ★★★★★

Distance : 7.4 miles

[Add to my Favorites](#)
[Map and Directions](#)

UNIV OF MI DIALYSIS CLINICS - LIVONIA

19900 HAGGERTY RD, SUITE 106
LIVONIA, MI 48152
(734) 432-7870

Rating : ★★★★★

Distance : 19.8 miles

[Add to my Favorites](#)
[Map and Directions](#)

[Learn more about star ratings](#)

[Learn why these characteristics and services are important.](#)



Star Ratings Dialysis Facility Compare Website

UNIV OF MI DIALYSIS - ADULT

1500 E MEDICAL CTR DR
D7605 UNIVERSITY HOSPITAL
ANN ARBOR, MI 48109
(734) 936-5656

Rating ⓘ: ★●●●●

Distance ⓘ: 4.3 miles

[Add to my Favorites](#)
[Map and Directions](#)

No

MICHIGAN DIALYSIS SERVICES OF ANN ARBOR

2850 S INDUSTRIAL HWY STE 100
ANN ARBOR, MI 48104
(734) 677-1490

Rating ⓘ: ★★☆☆●

Distance ⓘ: 7.4 miles

[Add to my Favorites](#)
[Map and Directions](#)

No

UNIV OF MI DIALYSIS CLINICS - LIVONIA

19900 HAGGERTY RD, SUITE 106
LIVONIA, MI 48152
(734) 432-7870

Rating ⓘ: ★★★★★

Distance ⓘ: 19.8 miles

[Add to my Favorites](#)
[Map and Directions](#)

The range is 1 to 5 stars. More stars mean better quality care compared to other facilities. A 1- or 2-star rating does not mean care is poor: it means that measured outcomes were below average compared to other facilities.

-By scrolling over the “i” next to rating, patients can view a brief description of how to interpret the Star Rating.



Star Ratings Display of Above and Below Average

The range is 1 to 5 stars. More stars mean better quality care compared to other facilities. A 1- or 2-star rating does not mean care is poor: it means that measured outcomes were below average compared to other facilities.

- Is the display of above or below average informative?
- Will patients know to scroll over the “i”?
- Is there a better way to display above or below average?



Educating Users About Star Ratings

How can I use the star ratings?

If you are new to dialysis, you and your doctor can talk about what the ratings mean and how you can use them with other information to help you decide where to go for treatment. We also suggest you visit a dialysis unit, look around, meet the staff, and **ask questions**.



How Star Ratings are Conveyed

- Are the descriptions of measures clear?
- How should the Star Rating quality information be used by consumers?
- Is the current DFC guidance on how to use star ratings helpful?
- How can we improve guidance on the DFC site?
- Recommendations for minimizing over-interpretation (misuse?)



Questions to Consider When Reviewing the Star Ratings on the DFC Website

- What key information should be included in any education materials developed for the Star Ratings?
- What is *easy* to interpret about the Star Ratings?
- What is *difficult* to interpret about the Star Ratings?
- Why would patients use the Star Ratings?
- Why would patients NOT use the Star Ratings?



Dialysis Facility Compare Website

www.medicare.gov/dialysisfacilitycompare/

Medicare.gov | Dialysis Facility Compare
The Official U.S. Government Site for Medicare

Home Share

Find a dialysis facility

A field with an asterisk (*) is required.

* Location
Example: 45602 or Lima, OH or Ohio

ZIP Code or City, State or State

Dialysis Facility Name (optional)
Full or Partial Dialysis Facility Name

Search

Need to find a dialysis facility while on vacation? We can help!

1 2 3 4 5

Spotlight

- New** Search for facilities based on their star rating. [Learn more.](#)
- Learn more about our [Quality Measures](#)
- Use these helpful resources when looking for a dialysis facility:
 - What to ask dialysis care providers
 - What to ask about dialysis centers

Additional Information

- Dialysis Facility Compare data last updated:** January 29, 2015
- Download the [Dialysis Facility Compare Database](#)
- Coming soon! A rate of hospital readmission measure will be added to Quality measures: hospitalizations and deaths.
- Get [Dialysis Facility Compare data archives](#)
- Dialysis facilities: Update your address, phone number, and other information
- Get End Stage Renal Disease (ESRD) Quality Incentive Program data.**

Tools & Tips

- Learn how Medicare covers dialysis services and supplies
- Report a concern or grievance/complaint about your dialysis or transplant care, or access to care
- Get tips for printing dialysis facility information
- Compare other providers and plans
 - Hospital Compare
 - Physician Compare
 - Home Health Compare
 - Nursing Home Compare
 - Medicare Plan Finder

12:32 PM 4/8/2015



Questions to Consider When Reviewing the DFC Website

- What is *easy* to interpret on the DFC site?
- What is *difficult* to interpret on the DFC site?
- Why would patients use the DFC site?
- Why would patients NOT use the DFC site?

Review of DFC Website Questions

- What key information should be included in any education materials developed for the star ratings?
- What is *easy* to interpret on the DFC site?
- What is *difficult* to interpret on the DFC site?
- Why would patients use the DFC site?
- Why would patients NOT use the DFC site?



DFC Star Rating TEP: Methodology Group Presentation

Presenter: Chris Harvey



DFC Star Ratings and Consumers

- Provides an easily recognizable way to compare facilities
- Offers additional information that consumers can use to make better informed decisions or ask questions, along with:
 - Visiting the facility and asking questions
 - Talking with a doctor
 - Looking up data on individual quality measures



Goals of the Methodology Group

- Review how the current methods combine the current DFC measures to create a summary rating
- Identify areas for improvement in the methodology
- Provide recommendations



Morning Presentation Overview

- DFC Measures and Star Rating Overview
- Measure Scoring
- Measure Weighting
- Star Categorization



Afternoon Presentation Overview

- Comparison of Methods discussed in presentation 1
- Missing Measure Values in Facilities
- Facility Size Adjustment
- 2 year comparisons
- Framework for adding new measures



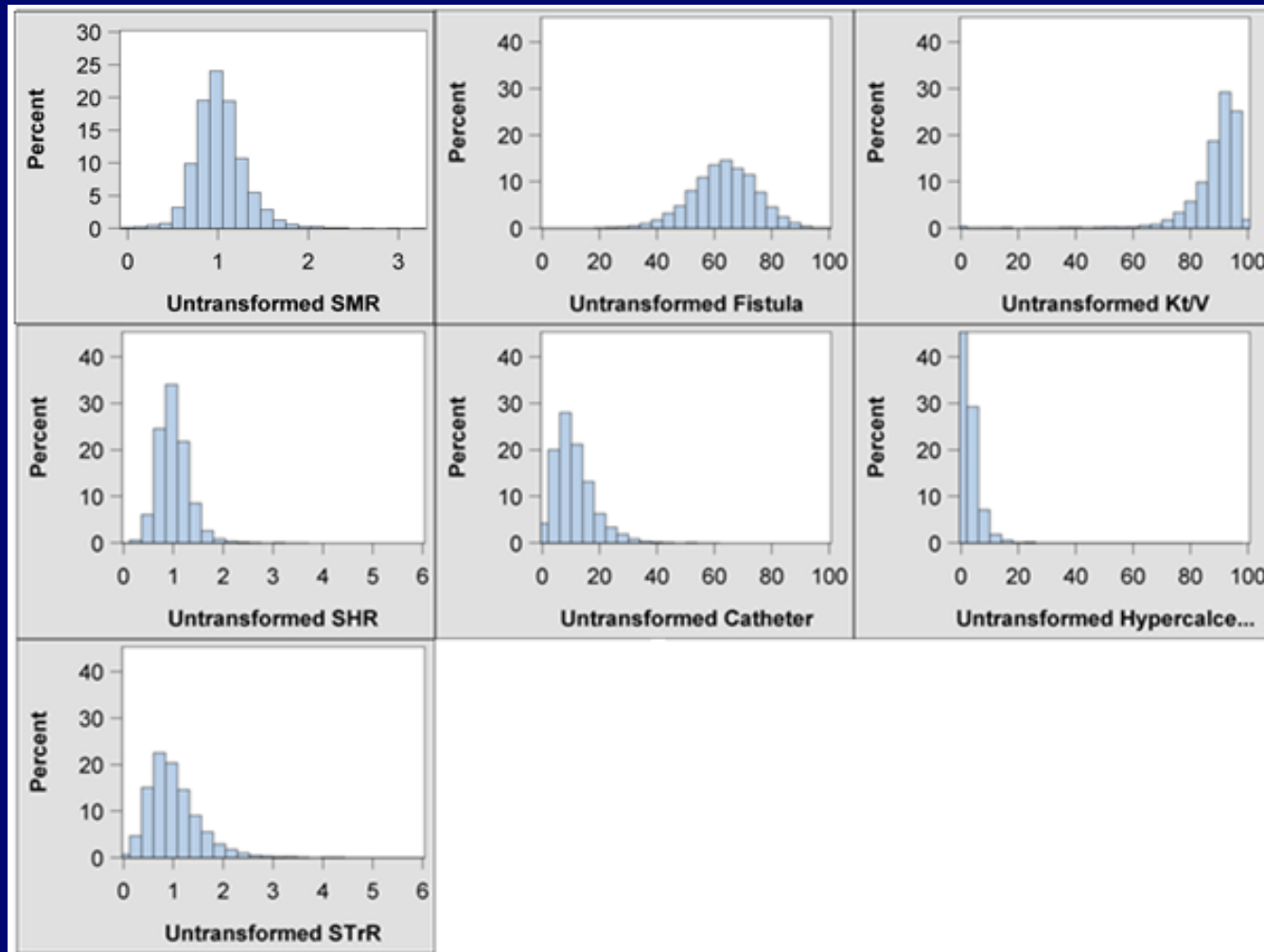
Methodology Group Presentation #1

Current DFC Star Rating

- Provides each facility a single rating to summarize 9 quality of care measures reported on DFC
- Current methods provide a solution to combine measures with different scales, distributions, and inter-correlations



Quality Measures and Distributions



3 Decisions to Make When Combining Measures for Overall Rating

Once the measures are decided, three major decisions will form a framework for creating the Star Rating

- **Decision 1:** Measure Scoring
- **Decision 2:** Measure Weighting
- **Decision 3:** Star Categorization



3 Decisions to Make When Combining Measures

Once the measures are decided, three major decisions will form a framework for creating the Star Rating

- **Decision 1: Measure Scoring**
- **Decision 2: Measure Weighting**
- **Decision 3: Star Categorization**



Decision 1: Some Measure Scoring Options

Minimal Transformation

- QMs are only adjusted in direction (so higher is better) and scale (ex. all measures range from 0 to 100)

Ranking Methods

- **Percentile Ranking** – ranking QMs on a uniform distribution between 0 and 100. (same number of facilities are given each value)
- **Probit Ranking** – ranking QMs on a normal distribution between 0 and 100. (more facilities are given a middle value than an extreme value)

Threshold Methods (giving measures their own groups or ratings)

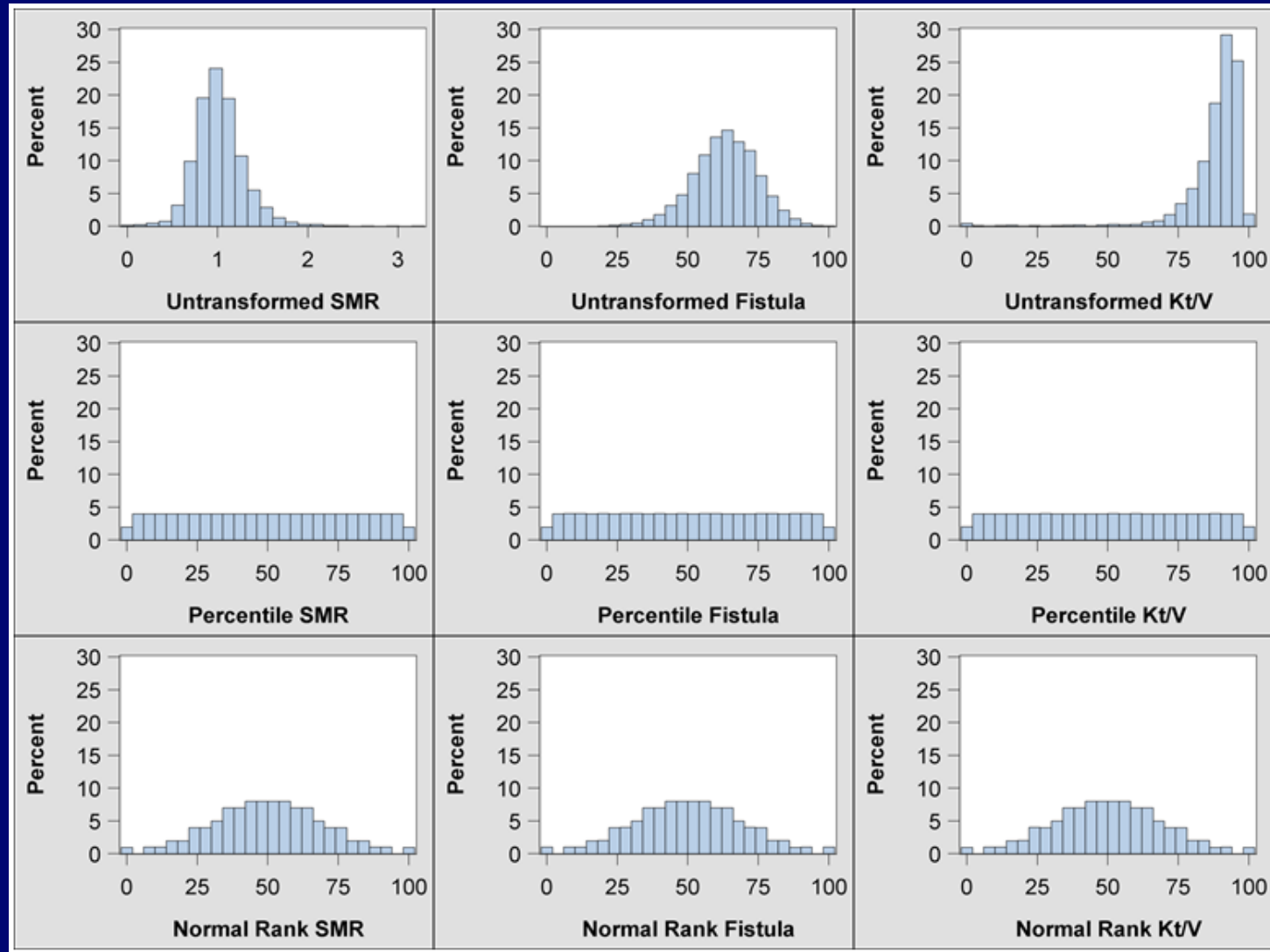
- **Clustering** – various methods that group QMs, so that groups contain values that are more similar to each other, and less similar to values in other groups.
- **Percentile Thresholds**– grouping QMs based on relationship to national average
- **Performance Thresholds** – grouping quality measures based on fixed values of the measure

Centering Methods

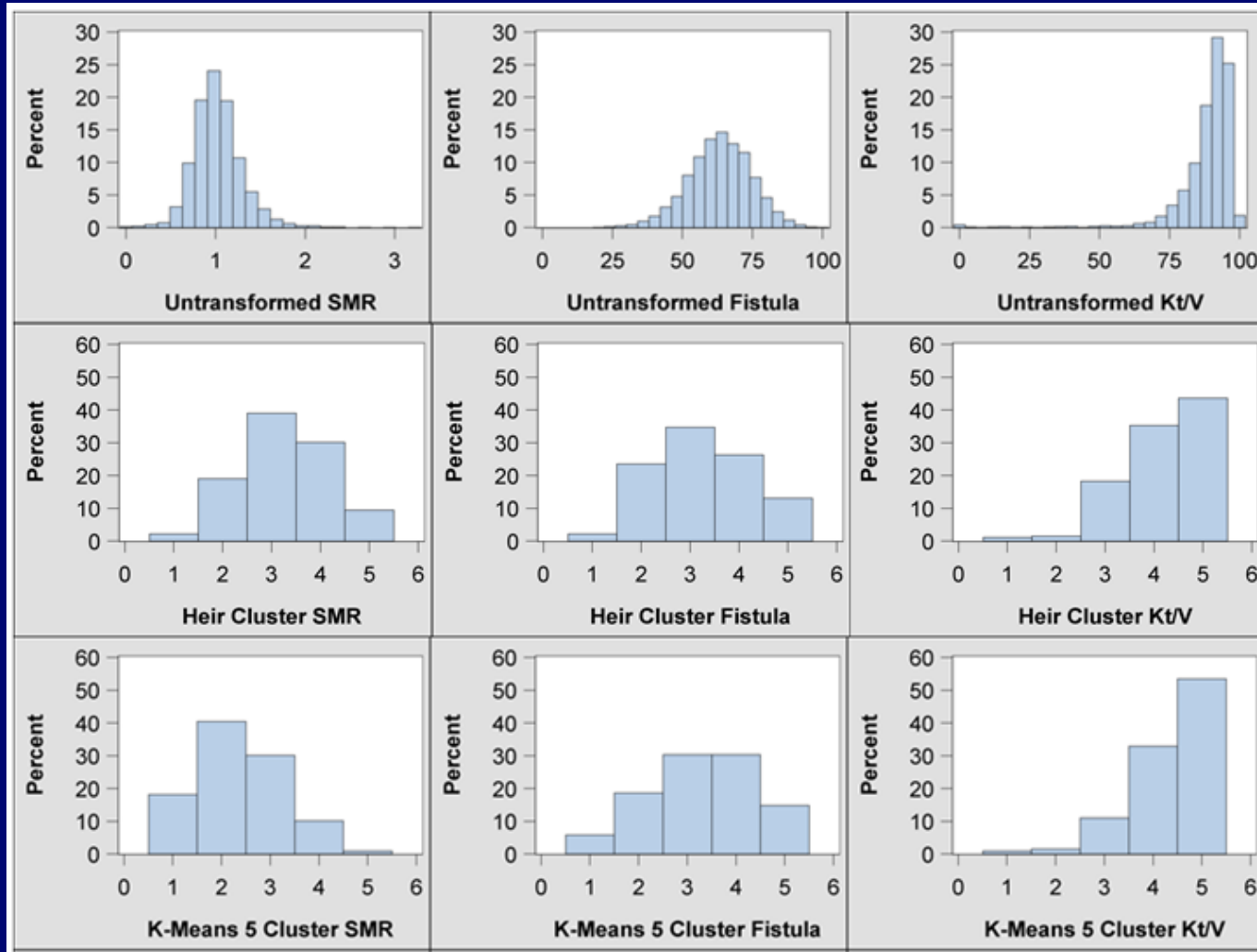
- **Z-Score** – how many standard deviations a QM value is away from the mean of that QM



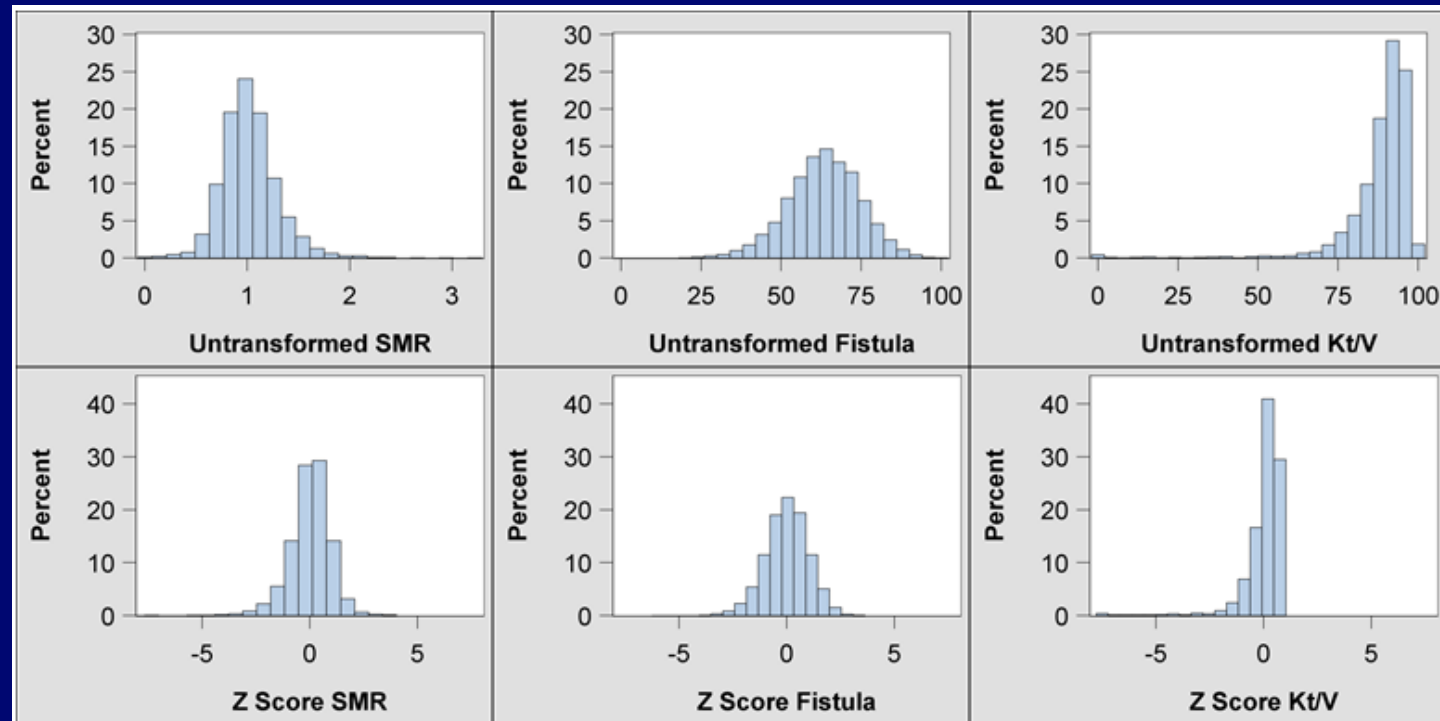
Decision 1: Visualizing Ranking Methods



Decision 1: Visualizing Clustering Methods



Decision 1: Visualizing Centering Methods



Summary

Decision 1- Measure Scoring

- Have to weigh advantages of probit ranking (controlling outliers, giving measures equal influence) and z-scores (preserving measure distribution)
- Categorization of measures at this stage could result in loss of information



3 Decisions to Make When Combining Measures

Once the measures are decided, three major decisions will form a framework for creating the Star Rating

- **Decision 1:** Measure Scoring
- **Decision 2:** Measure Weighting
- **Decision 3:** Star Categorization



Decision 2: Some Measure Weighting Options

- Equal Weighting
- Importance Weighting
 - No established consensus
- Adjusting for Redundancy
 - Groups of Measures are formed based on correlations with the aid of factor analysis, and groups are equally weighted



Spearman Correlation of Measures

(Groupings from Factor Analysis Highlighted)

| Measures | STrR | SHR | SMR | Kt/V | Hypercal | Fistula | Catheter |
|----------|------|-------------|-------------|------|-------------|---------|-------------|
| STrR | 1.00 | 0.40 | 0.21 | 0.08 | 0.00 | 0.11 | 0.15 |
| SHR | | 1.00 | 0.26 | 0.11 | 0.01 | 0.13 | 0.19 |
| SMR | | | 1.00 | 0.08 | 0.05 | 0.17 | 0.11 |
| Kt/V | | | | 1.00 | 0.19 | 0.06 | 0.13 |
| Hypercal | | | | | 1.00 | 0.09 | 0.05 |
| Fistula | | | | | | 1.00 | 0.45 |
| Catheter | | | | | | | 1.00 |



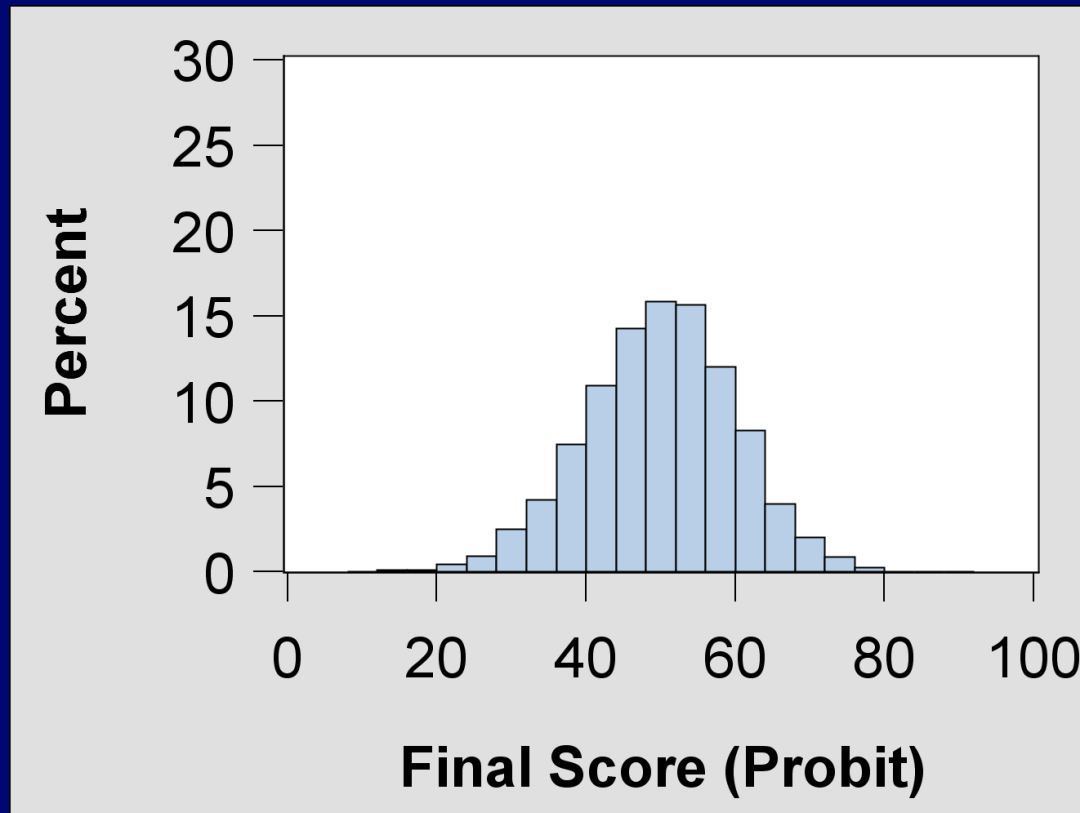
Summary

Decision 2- Measure Weighting

- Current methods creates domains of measures based on correlations
- Measures within domains are equally weighted to give a domain score
- Domains are equally weighted to give each facility a final score



Facility Final Scores



3 Decisions to Make When Combining Measures

Once the measures are decided, three major decisions will form a framework for creating the Star Rating

- **Decision 1:** Measure Scoring
- **Decision 2:** Measure Weighting
- **Decision 3:** Star Categorization



Decision 3: Various Star Categorization Options

- **Percentile Thresholds**
 - fix the annual proportion of facilities in each star rating category
- **Quality Thresholds**
 - fix a final facility scores in each rating or require certain scores on each measure/ group of measures to attain rating
- **Final Score Clustering**
 - group final scores with statistical clustering, so that groups contain values that are more similar to each other, and less similar to values in other groups.
- **Average QM Star Ratings**
 - rounding star ratings created for individual measures



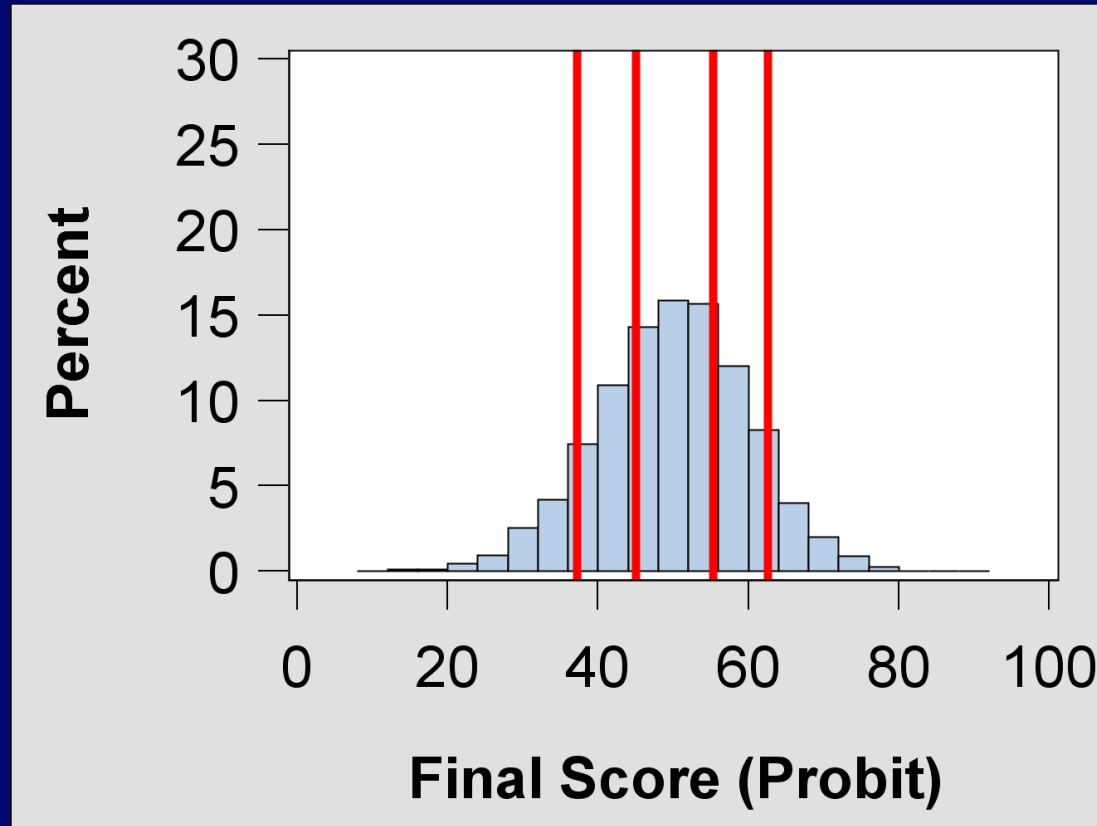
Decision 3: Various Star Categorization Options

- **Percentile Thresholds**
 - We chose fixed deciles:
 - 10% 1-Star and 5-Star, 20% 2-Star and 4-Star, 40% 3-Star
 - Fixing top and bottom performers may be problematic if distribution of facility scores changes over time
- **Quality Thresholds**
 - Fixing final scores difficult because standardized measures based relative to other facilities for that year
 - Fixing measure values cut-offs which essentially groups individual measures results in loss of information
- **Final Score Clustering**
 - Different clustering methods can give different results
 - Outliers can form own clusters
- **Average QM Star Ratings**
 - Fixing measure value of grouping cut-offs results in loss of information



Percentile Thresholds:

10% 5-Star, 20% 4-Star, 40% 3-Star, 20% 2-Star, and 10% 1-Star



Summary: DFC Star Rating

Decision 1: Rank measures with probit ranking

Decision 2: Create domains of correlated measures with aid factor analysis and equally weight groups

Decision 3: Use percentiles for ratings :

- 10% 1-star, 20% 2-star, 40% 3-star, 20% 4-star, 10% 5- star



Questions ?

Methodology Group Presentation #2

Presenter: Chris Harvey

Presentation Overview

- Comparison of Methods
- Missing Measure Values in Facilities
- Facility Size Adjustment
- 2 year comparisons
- Framework for adding new measures
- Recommendations from the Community
- Summary Conclusion



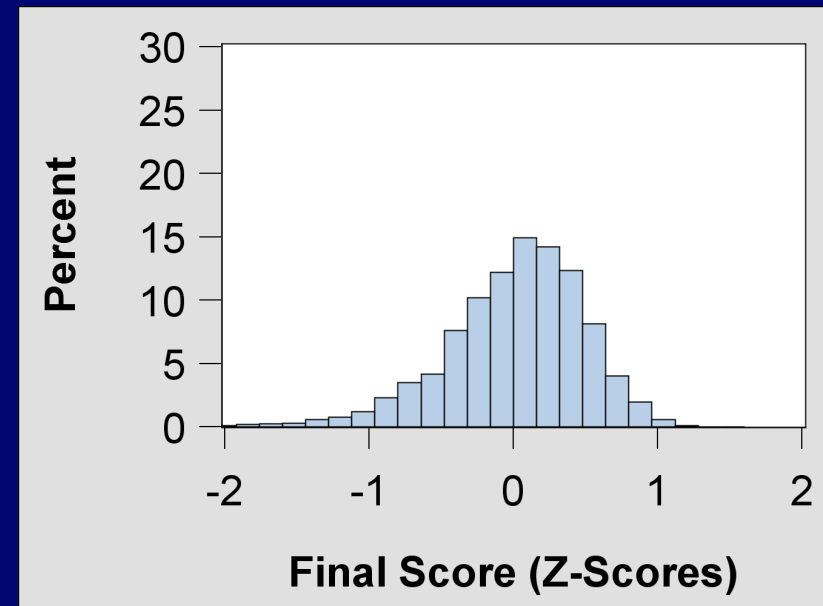
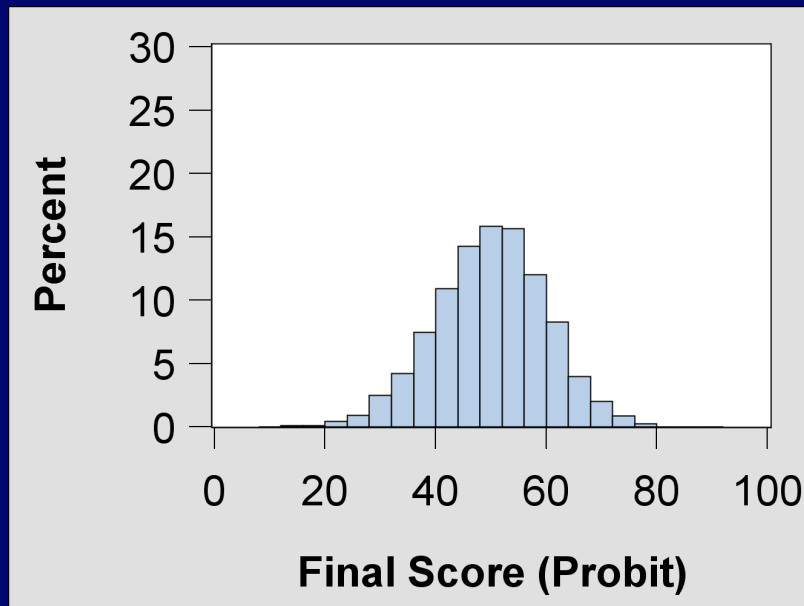
Comparison of Methods:

- Considering Z-scores in place of Probit Ranks for measure transformation



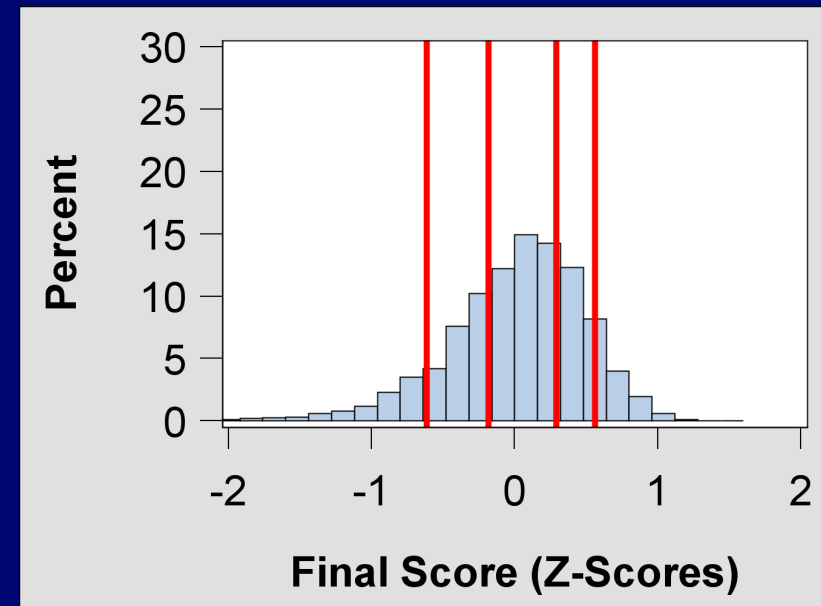
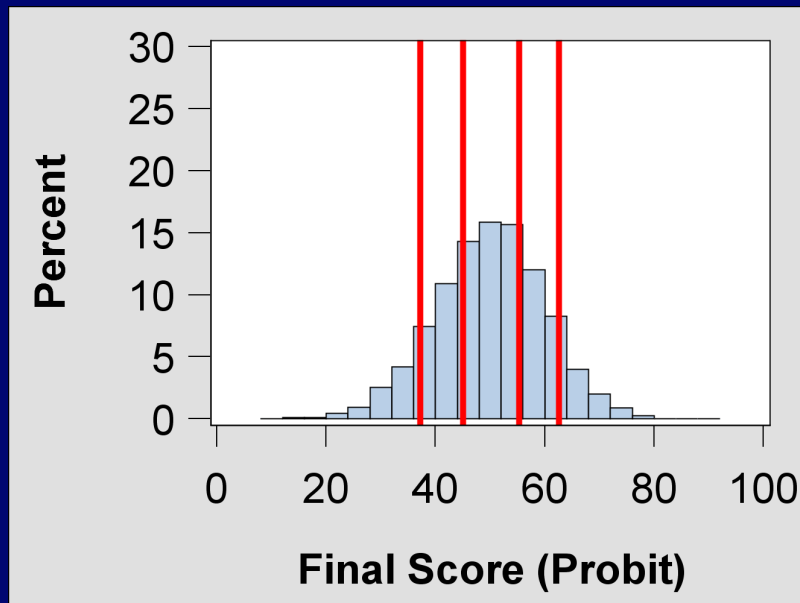
Distribution of Final Scores:

(Probit Scored Measures Vs. Z-Scored Measures)



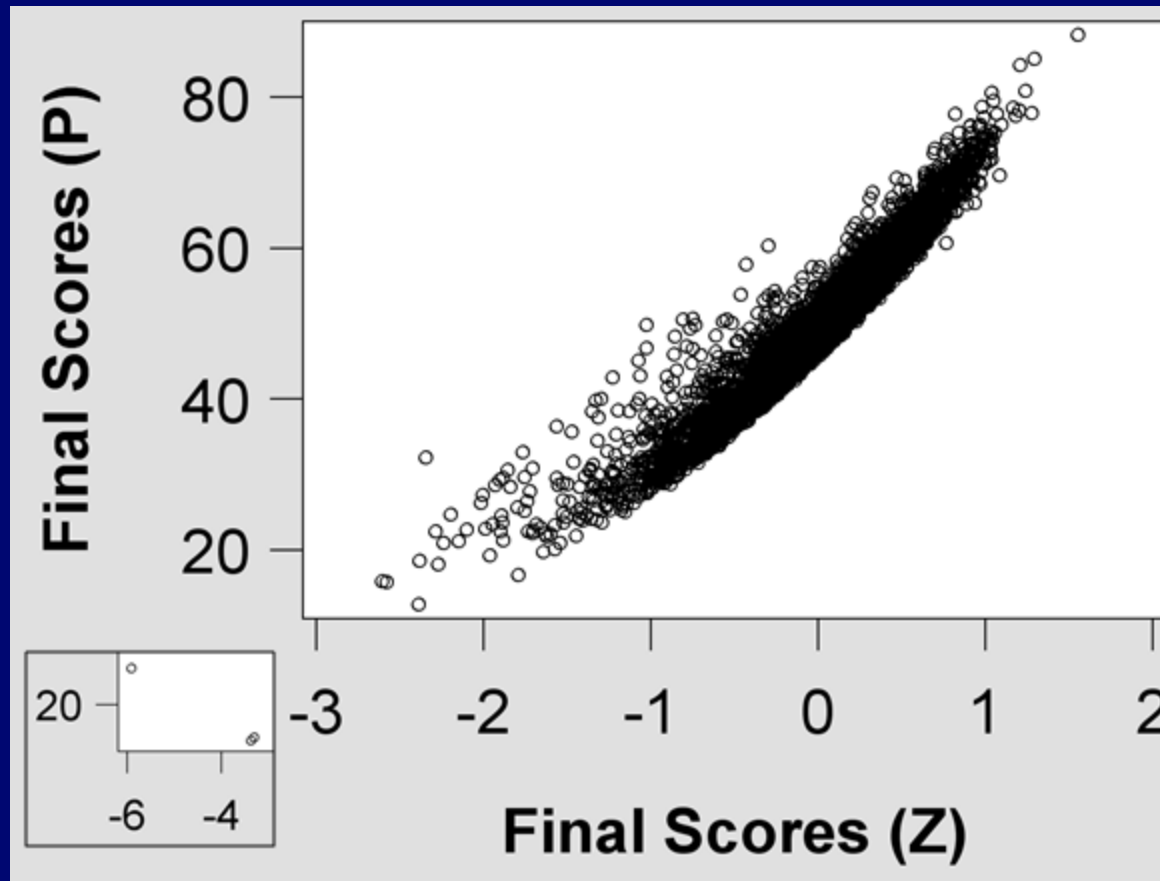
Distribution of Final Scores and Visualization of Star Rating Categories

(Probit Scored Measures Vs. Z-Scored Measures with fixed deciles)



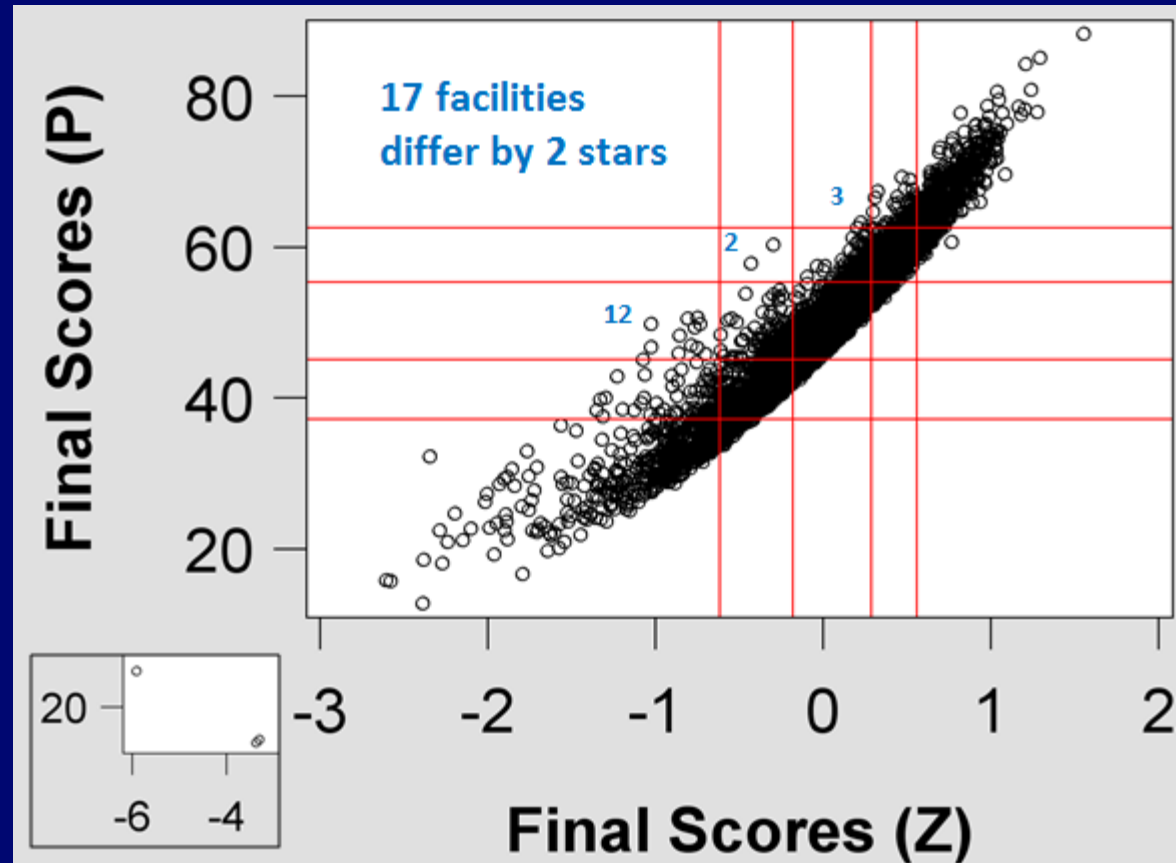
Distribution of Final Scores:

(Probit Scored Measures Vs. Z-Scored Measures with fixed deciles)



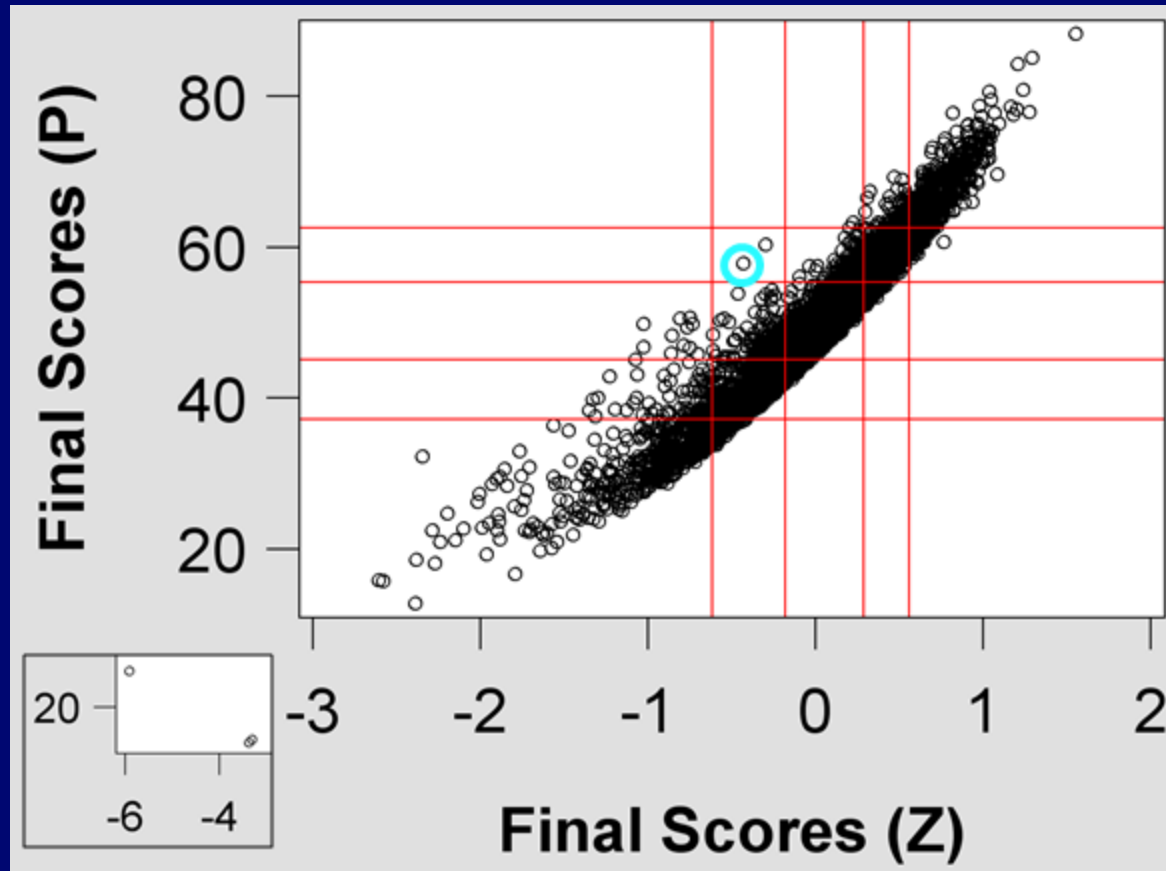
Distribution of Final Scores:

(Probit Scored Measures Vs. Z-Scored Measures with fixed deciles)



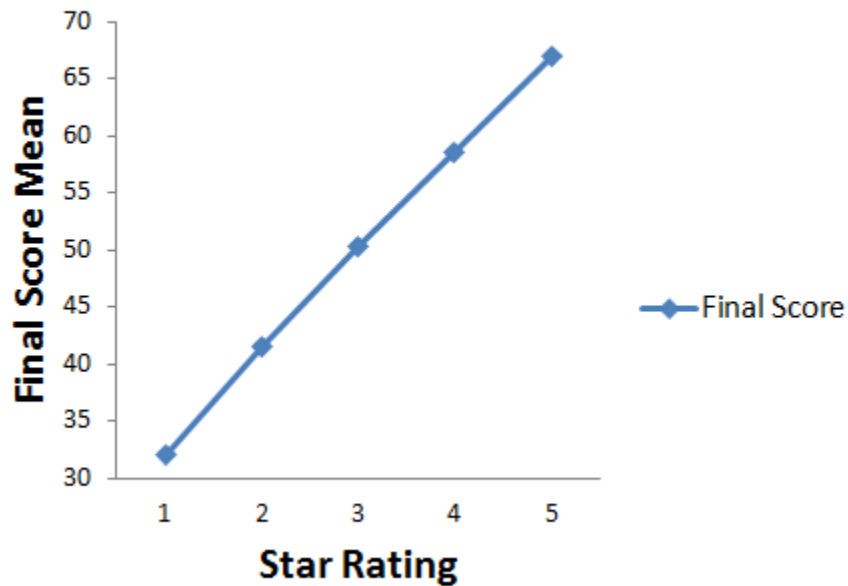
Distribution of Final Scores:

(Probit Scored Measures Vs. Z-Scored Measures with fixed deciles)

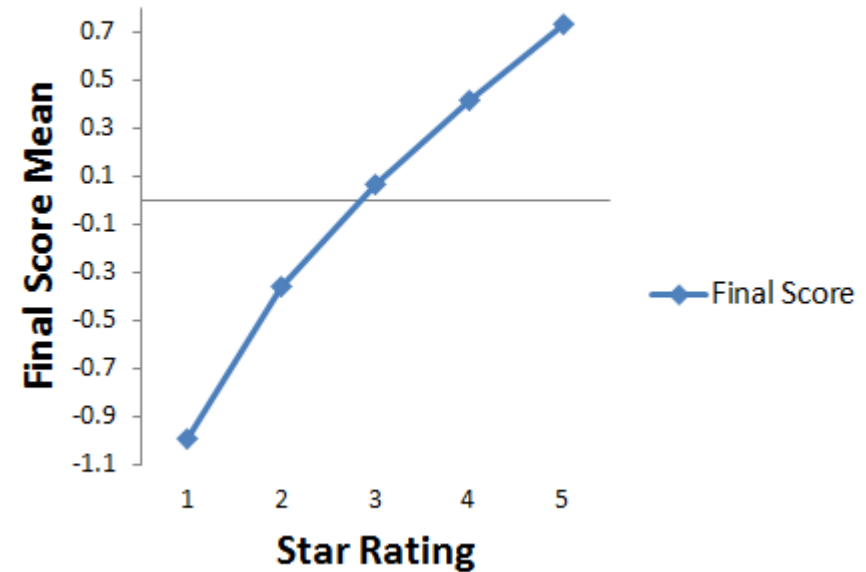


Mean Final Scores in Adjacent Tiers in DFC Ratings

Probit Ranked Measures

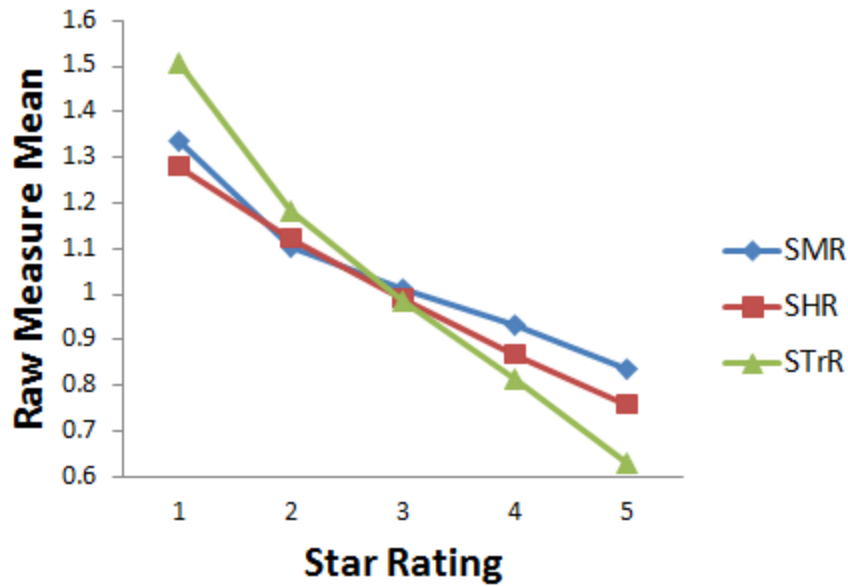


Z-Scored Measures

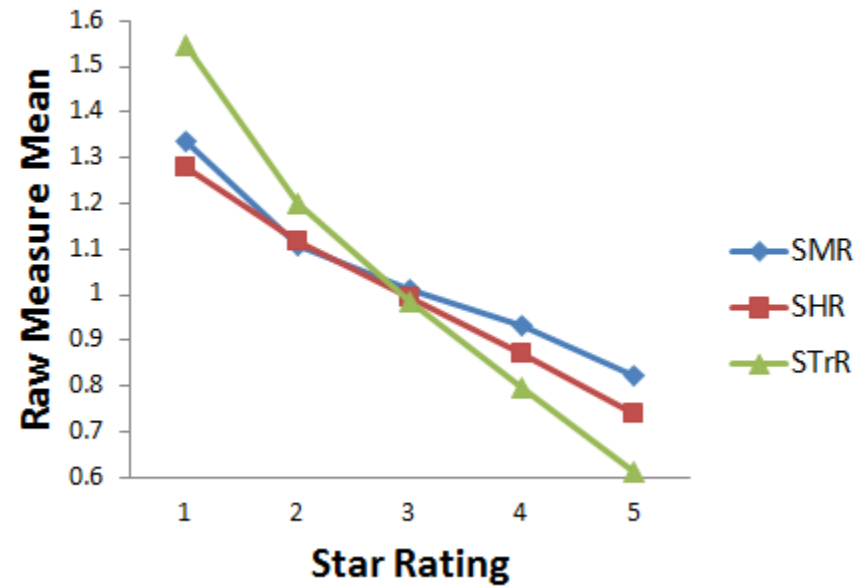


Standardized Ratio Measures by DFC Star Rating Tiers

Probit Ranked Measure

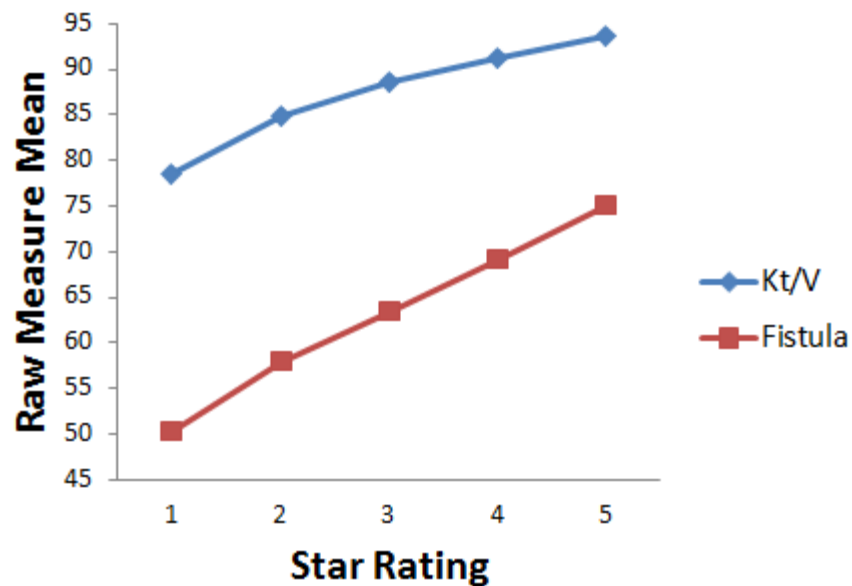


Z-Scored Measures

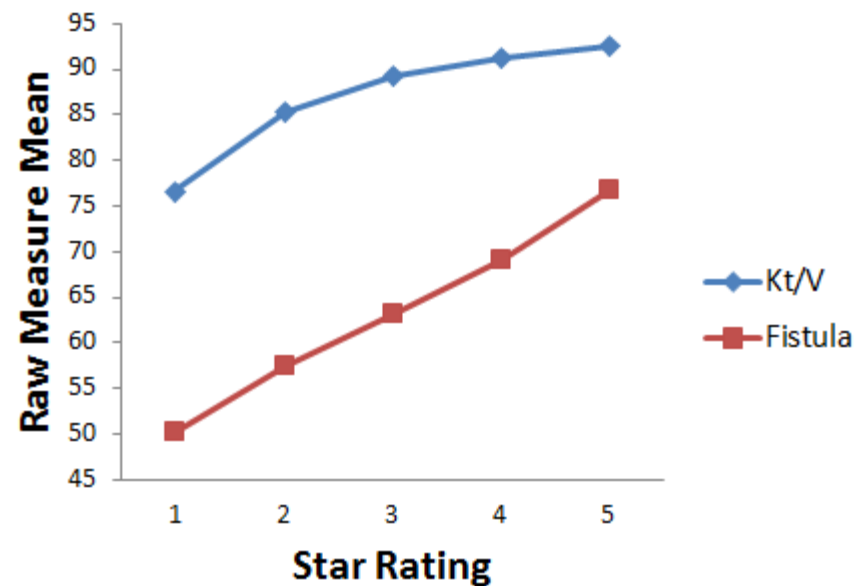


Percentage Measures by DFC Star Rating Tiers (Higher is Better)

Probit Ranked Measures

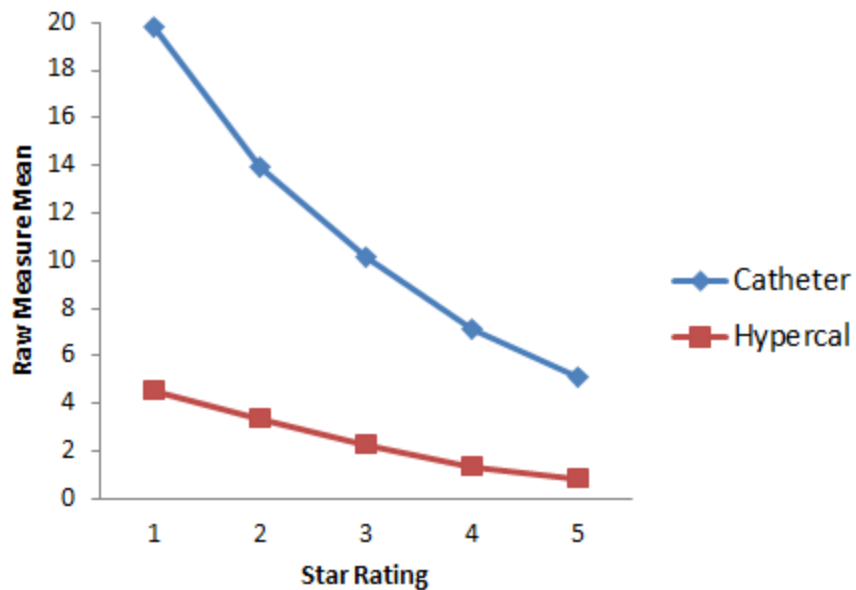


Z-Scored Measures

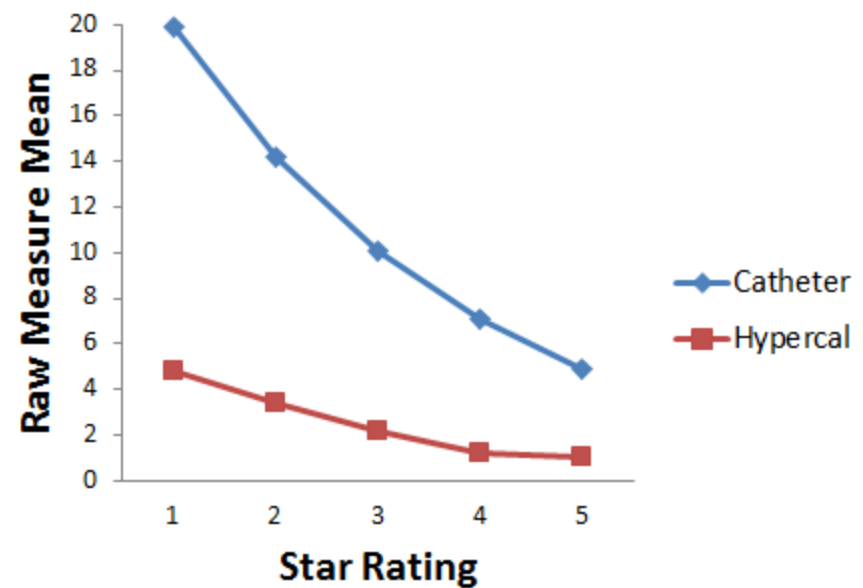


Percentage Measures by DFC Star Rating Tiers (Lower is Better)

Probit Ranked Measures



Z-Scored Measures



Comparison of Methods:

- Considering statistical clustering of Final Scores rather than fixed deciles



Clustering DFC Final Score (K-Means)

| Cluster (low to high) | Probit Ranked Measures | | Z-Scored Measures | |
|-----------------------------|------------------------------|-----|----------------------|-----|
| | N | % | N | % |
| 1 | 530 | 10% | 1 | 0% |
| 2 | 1275 | 24% | 324 | 6% |
| 3 | 1620 | 30% | 1348 | 24% |
| 4 | 1398 | 25% | 2262 | 40% |
| 5 | 594 | 11% | 1700 | 30% |



Sensitivity of the Rating

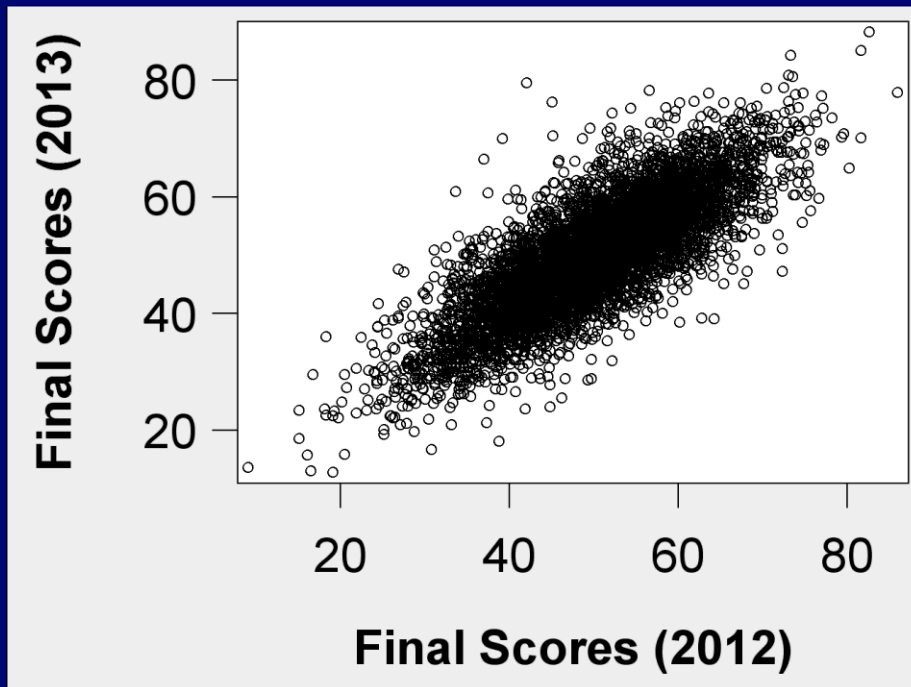
- If measures were completely random, probability of having the same star rating two years in a row would be 26%

| Star Rating | Probability of same rating for 2 years by chance |
|-------------|--|
| 1 | 0.01 |
| 2 | 0.04 |
| 3 | 0.16 |
| 4 | 0.04 |
| 5 | 0.01 |
| SUM | 0.26 |

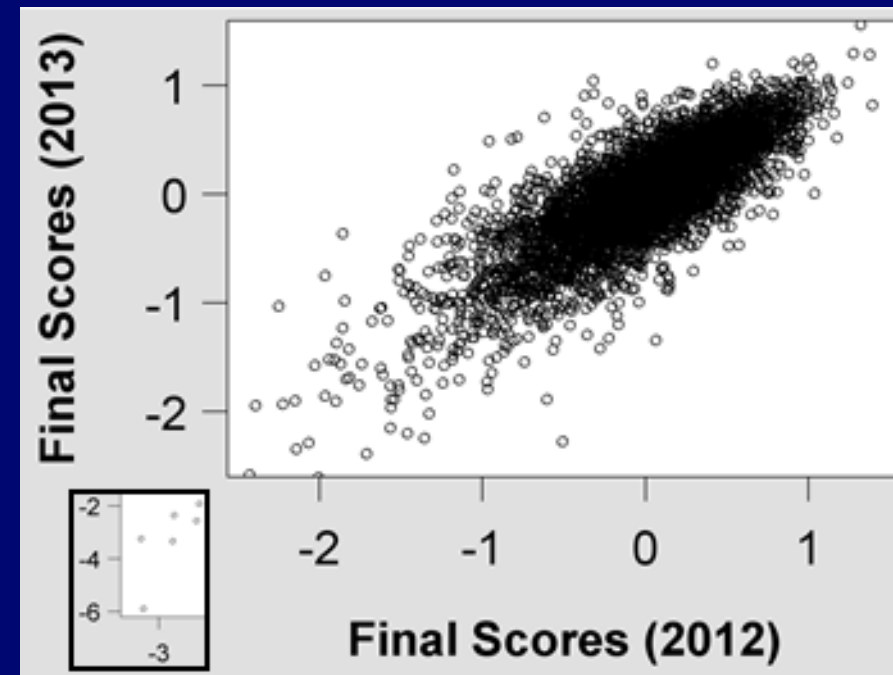


Sensitivity of Rating: 2 year Comparison

Using Probit Ranked Measures

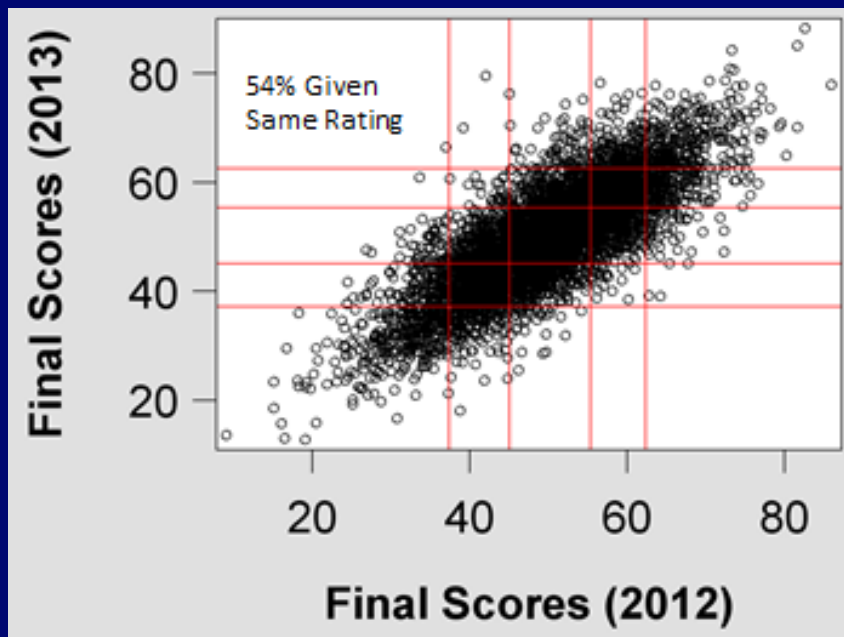


Using Z-Scored Measures

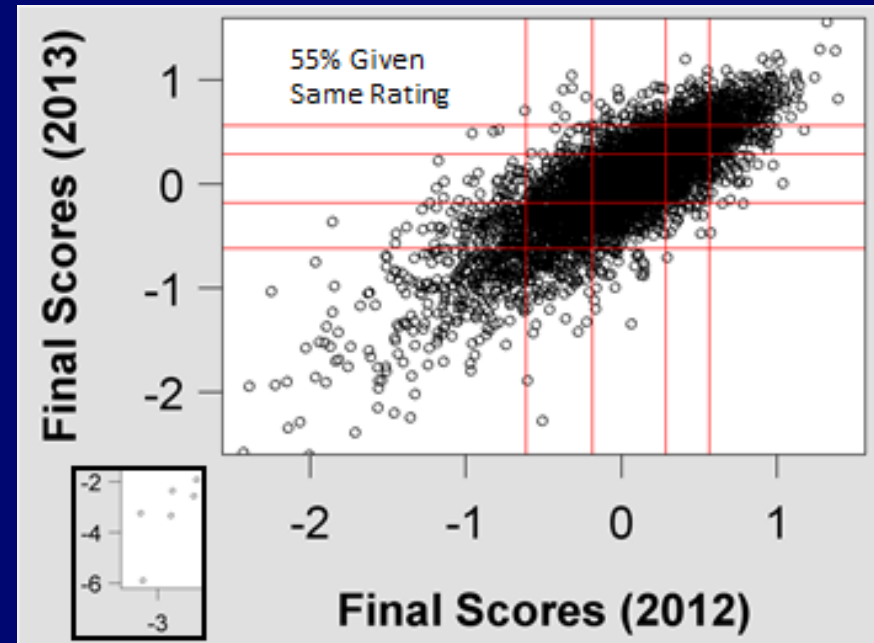


Sensitivity of Rating: 2 year Comparison

Using Probit Ranked Measures



Using Z-Scored Measures



Sensitivity of Clustering: 2 year comparison

| Measure Scoring | Probit Rank | | Z-Score | | |
|--|-------------|------|---------|------|-----|
| | 2012 | 2013 | 2012 | 2013 | |
| KMEANS Clusters (low to high) | 1 | 10% | 10% | 2% | 0% |
| | 2 | 24% | 24% | 12% | 6% |
| | 3 | 30% | 30% | 28% | 24% |
| | 4 | 26% | 25% | 36% | 40% |
| | 5 | 11% | 11% | 23% | 30% |
| Hierarchical Clusters (low to high) | 1 | 11% | 16% | 1% | 1% |
| | 2 | 28% | 24% | 11% | 12% |
| | 3 | 41% | 20% | 30% | 17% |
| | 4 | 16% | 21% | 26% | 36% |
| | 5 | 4% | 19% | 33% | 34% |



Other Suggestions from the Community

- Create domain/measure thresholds necessary to obtain 1-Star or 5-Star rating.
- Scoring measures based on confidence intervals
- Shifting Star Rating Cutoffs based on confidence intervals



Other Suggestions from the Community

- Create domain/measure thresholds necessary to obtain 1-Star or 5-Star rating.
- Scoring measures based on confidence intervals
- Shifting Star Rating Cutoffs based on confidence intervals



Domain Thresholds

- Should facilities score above a specific score on each domain (or measure) to receive 5-stars?
- Should facilities score below a specific score on each domain (or measure) to receive 1-star?



Domain/Measure Thresholds

- Advantages
 - ensures 5-star facilities are above average across the board
 - ensures 1-star facilities are below average across the board
- Disadvantages
 - a much above average facility may not be recognized as 5-star due to performance on one measure/domain
 - a much below average facility may not be recognized as 1-star due to performance on one measure/domain



Other Suggestions from the Community

- Create domain thresholds necessary to obtain 1-Star or 5-Star rating.
- Scoring measures based on confidence intervals
- Shifting Star Rating Cutoffs based on confidence intervals



Scoring measures based on confidence intervals ⁵⁰

- Motivation
 - DFC reports Standardized Measures as “better than expected”, as expected, and “worse than expected” based on the 95% Confidence interval
 - Scoring measures discretely based on these intervals a suggestion to account for uncertainty based on facility size.

Other Suggestions from the Community

- Create domain thresholds necessary to obtain 1-Star or 5-Star rating.
- Scoring measures based on confidence intervals
- **Shifting Star Rating Cutoffs based on confidence intervals**



Shifting Star Rating Cutoffs based on confidence intervals

- Was suggested that if we created star ratings based on domain thresholds, these cutoffs would have uncertainty
- suggest making a confidence interval around the cutoff and choosing the lower bound as the new cutoff
- Advantage: avoid misclassifying facilities as below average

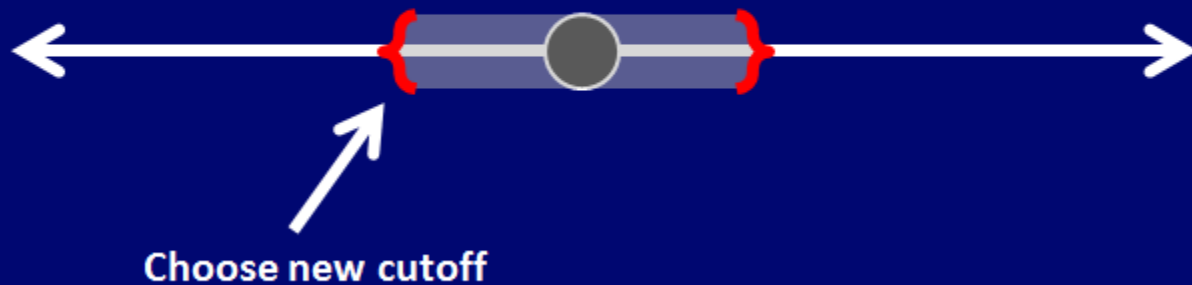
Shifting Star Rating Cutoffs based on confidence intervals



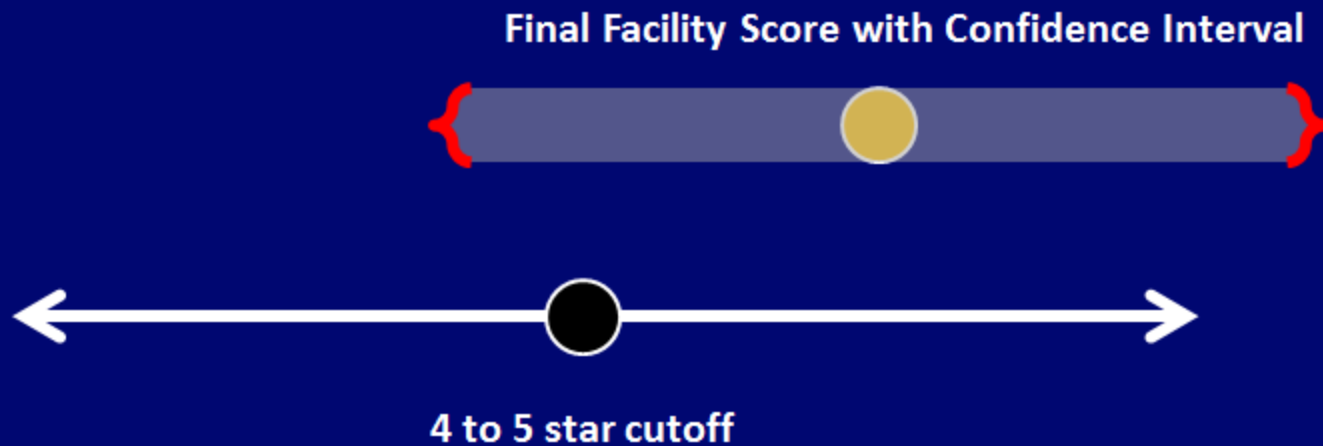
Shifting Star Rating Cutoffs based on confidence intervals



Shifting Star Rating Cutoffs based on confidence intervals



Shifting Star Rating Cutoffs based on confidence intervals



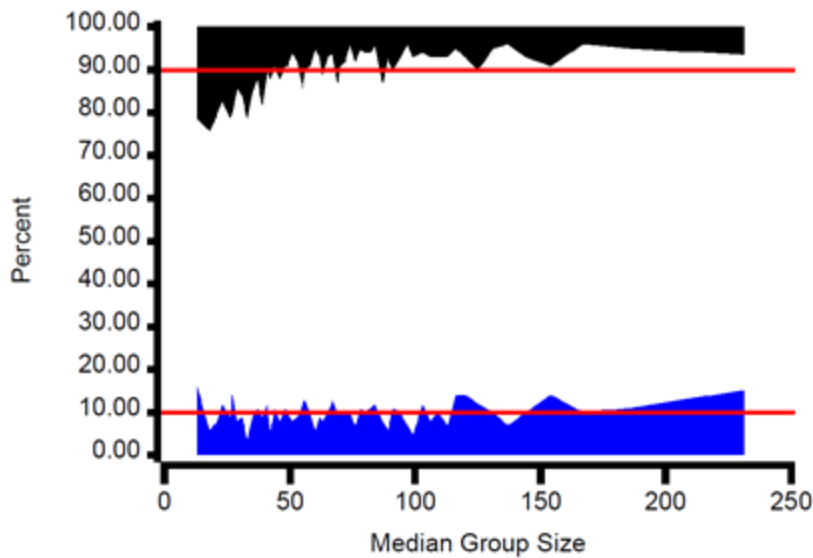
Facility Uncertainty Adjustment: Current Employed Methods

- Facilities missing a domain are not given a rating
- If no domains are missing, missing measures are imputed with the national average rank (probit rank of 50)
 - With less information, we shrink towards average
- Should a similar adjustment be used on small facilities?

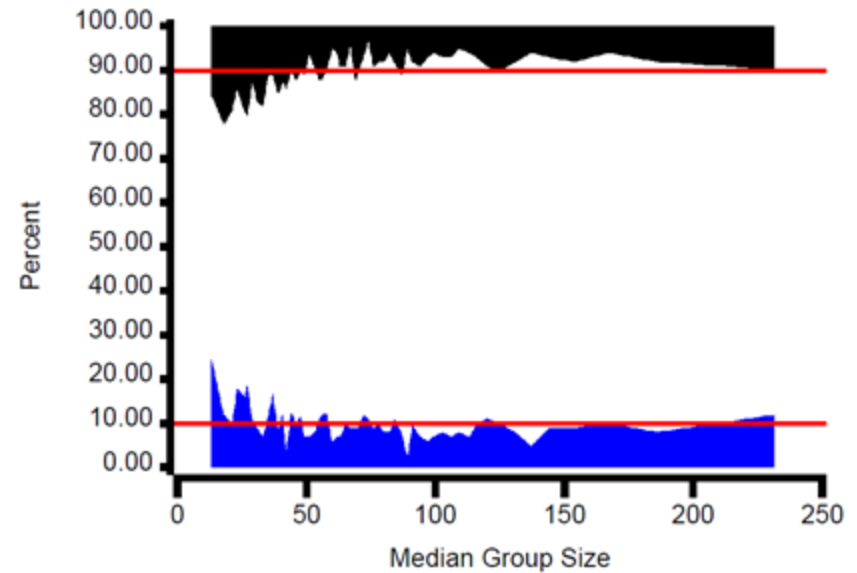


Distribution of Star Ratings by facility size

Probit Ranked Measures



Z-Scored Measures



■ % 5-Star

■ % 1-Star



Framework for adding more measures

Framework for adding more measures

- Methodology needs to be flexible to accommodate the addition or removal of measures
- Reconsider measure domains with each new iteration (including when new measures are added)
- Should we ever consider using a domain with a single measure?
 - Could result in many facilities not receiving a rating
 - Could give that measure too much influence on the rating
- Some methods may result in more drastic changes with addition of new measures



Summary of Topics Covered

- DFC Measures and Star Rating Overview
- Measure Scoring
- Measure Weighting
- Star Categorization
- Comparison of Methods discussed in presentation 1
- 2 year comparisons
- Missing Measure Values in Facilities
- Facility Size Adjustment
- Framework for adding new measures



Questions ?



Patient Workgroup- Day 1

- For risk-adjusted measures (SMR, SHR, STrR), SES adjustment is an important factor. Some in favor of using; others feel that the potential benefits and risks need exploration
- Methodology used is difficult to understand for intended users
- Methodology used is not consistent with public ratings based on on-line ratings. Needs wordsmithing to reflect the difference between the results presented here using stars and the fundamental difference with other sites use of stars for consumer reported satisfaction.



Patient Workgroup- Day 1

- In general, overall perceived relevance of DFC measures for patients/consumers is low
- Setting an established standard is preferred over relative or “curved” rankings when possible
 - Include multiple standards or thresholds to indicate partial achievement
 - Include opportunity for facility comments/explanations
- Consistency across ESRD programs- discussion mixed with some members clearly supportive and others less certain

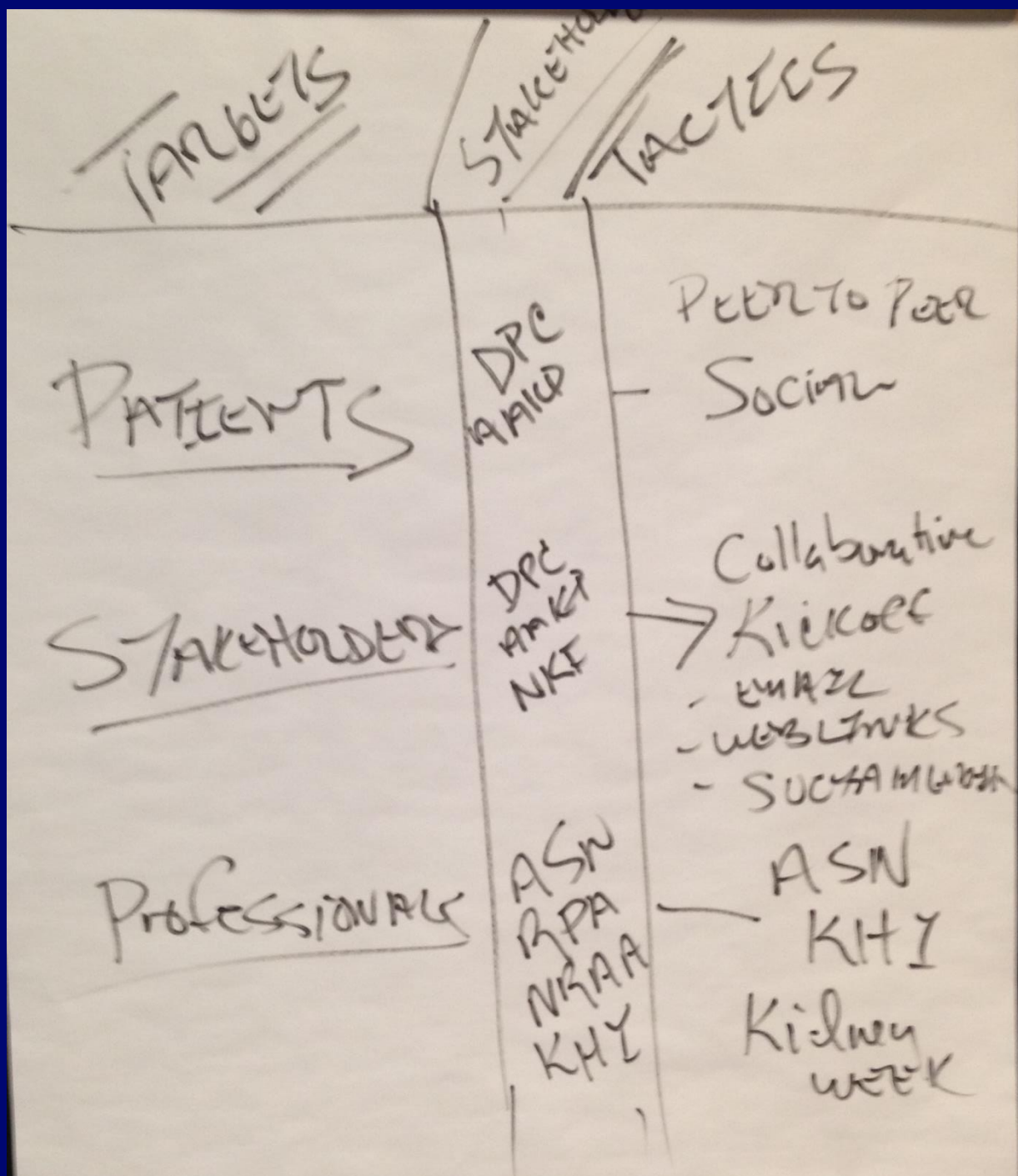


Patient Workgroup- Topic Area

Recommendations

- Consensus for Addition (greater concern that additions are needed)
 - Safety Outcomes (e.g. Infection outcomes, falls, medication errors, adverse event reporting, etc.)
 - Patient-reported outcomes (quality of life, patient-assessed quality of care, assessment of staff training/performance, promoting modality choice, staff responsiveness for patient concerns, others?)
 - Broader consumer testing of current measures in the Star Rating Program is needed to assess relevance to consumers (SMR, SHR, vascular access, adequacy).
- When asked about possible measure retirement, Patient Workgroup identified Hypercalcemia.





Rollout

STAKEHOLDERS
[FEEDBACK]



BETA TEST
[PARTNERS]



PRE-ROLLOUT Briefing
[STAKEHOLDERS]
EMBARGOED brief
↓ (MEDIA)

ROLLOUT ————— STAKEHOLDERS



Methodology Group

Summary Recommendations

Major Discussion Points

- Scoring
- Weighting
- Categorization

- Missing Data (Imputation)
- Uncertainty
- Data Presentation

Scoring

- Issue 1: How to combine measures that are on different scales
- Issue 2: How to reduce the effects of imprecise measures (e.g., for small facilities)

Scoring

Current method: Rank probit scoring

- Advantages:
 - Strong statistical properties
 - Addresses both issues above
 - Maintains relative standings
- Disadvantages:
 - Does not retain information on actual spacing in the original scale of measurement
 - May appear less clinically transparent

Scoring

Alternative: Z-scores

- Advantages:
 - Simple change of original units to standard deviation units
 - Retains information on the original spacing
 - Addresses issue 1
- Disadvantages:
 - Does not address issue 2
 - Gives more weight to extreme values when combining measures

Weighting

- Goal: To combine measures into a single measure that can be transformed into stars
- Stage 1: Combine measures that are effectively measuring the same thing (Factor Analysis)
 - Avoids letting one concept dominate the overall ratings because it has multiple measures.
 - Must be redone when measures are added or deleted.
 - Data-driven methods to create factors may lack direct clinical interpretability. Instead, clinically-determined groupings could be used.
 - The 7 current measures end up effectively equally-weighted, so this step does not current affect the final scoring. However, this could change with additional or fewer measures.

Weighting

- Stage 2: Combining domains into a single score
 - Currently: Average measures within domains; equally weight domains in final score.
 - Deemed both methodologically sound and transparent to the consumer.

Categorization into a Single Star Rating

- Currently: Stars are defined based on fixed percentiles

1* : 0 - 10 percentile

2* : 10 - 30

3* : 30 - 70

4* : 70 - 90

5* : 90 - 100

- Essentially “grading on a curve” - methodologically appropriate for goal of relative ratings.
- Three * represents the national average, and other categories are relative to that.
- Is not anchored to the clinical interpretations of the actual underlying measures.

Recommendations

- Need to anchor the stars in clinically meaningful terms.
 - Average scores for each measure and combined domains can be reported numerically and/or graphically for each star rating
 - Actual facility-level measures and associated percentiles, measures of uncertainty, etc. should also be reported on the web site.
- Need to impute missing values in a more informative way.
 - Current approach: no final score if all measures on a domain are missing; national average is assumed for other missing values.
 - The average facility size is around 60-70. Small facilities are more likely to have missing values.
 - Recommendation: use more facility level information to impute.

Recommendations

- Need to present information on uncertainty in ratings
 - Proposed several possible approaches, which KECC will investigate for performance properties