Technical Notes on Dialysis Facility Compare Star Rating System

Contents

Introduction 2

Overview of Measures 2

Quality Measures Used in Star Rating Calculation 2

Developing Quality Measure Domains 3

Analytic Approach 3

Standardization of Measures 3

Factor Analysis 5

Quality Measure Domains 6

Overall Star Rating for each Facility 6

Conclusions 8

References 8

Introduction

CMS, through a contract with UM-KECC, developed a Star Quality Rating System to rate the quality of care provided by dialysis facilities. The goal of the Star Rating System is to provide patients, their families, and caregivers information that they can use to easily compare dialysis facilities as well as be aware of areas of care delivery where the quality of care is rated lower. This document describes an overall quality rating system that gives each facility a rating between one and five stars. Facilities with five stars are considered to deliver much above average quality of care and those with one star are considered to deliver care that is rated much below average quality, compared to other dialysis facilities in the nation.

Overview of Measures

A set of Dialysis Facility Compare (DFC) Quality Measures (QMs) has been developed over the past 10 years. These are currently implemented on DFC and are used to rate the quality of care at all Medicare certified facilities. We used nine of the eleven QMs reported on the Medicare DFC website in the algorithm to determine the Star Rating for facilities using January 2013 release data.[1] The URR (measure of dialysis adequacy) and hemoglobin (measure of anemia management) measures were not used in this rating system because most patients achieve the goal values (national averages are 99% and 0% respectively) resulting in very little variability across facilities. Additionally, the three QMs measuring Kt/V levels are combined resulting in seven final measures used to rate facilities.

Quality Measures Used in Star Rating Calculation

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

There are currently three separate measures that report on a facility's achievement of removing enough wastes from the blood using Kt/V measurements for different types of patients, either based on modality, or for pediatric patients with HD as their modality. These are, respectively, measures for adult HD, adult PD, and pediatric HD patients. However, many facilities do not have peritoneal dialysis

[1] SMR is based on previous 4 years of data. All other measures are based on previous year of data.

patients and/or have few to no pediatric hemodialysis patients. To improve the ability to compare facilities with these different patient types, these three Kt/V measurements were combined into one measure. The percentage of patients that achieve Kt/V greater than the specified thresholds for each of the three respective patient types (adult PD patients, adult HD patients, and pediatric HD patients), was weighted based on the number of patient-months of data available. The resulting pooled measure (all Kt/V) represents the percentage of total dialysis patients who had enough wastes removed from their blood (Kt/V greater than or equal to specified threshold). After these measures were combined, there were seven final measures used to rate the dialysis facilities.

Developing Quality Measure Domains

Analytic Approach

A straight forward way of constructing an overall rating would be to use the un-weighted average of the seven final QMs. The correlation structure of the QMs (Table 1) reveal some measures are more correlated than with the others, which might cause issues with the equal weighting. Specifically, if some correlated QMs measure a similar aspect of quality of a facility and fewer QMs measure a different quality of a facility, equal weighting would artificially count the preceding quality as more important. We addressed this problem by grouping QMs in an unbiased manner by using factor analysis.

Factor analysis is a method for reducing a set of variables into groupings or latent factors that measure similar qualities based on the observed covariance structure (Johnson & Wichern, 2007). By grouping QMs into different domains, we can develop a final score based on equal weights of these latent factors which can be used to partition facilities into 5 different "star" levels. Equal weighting of these domains rather than the individual QMs avoids overweighting large groups of associated measures.

Standardization of Measures

The DFC QMs are noticeably different in distributions as well as scales. In order to make measures comparable across facilities and to reduce the impact of few possible outliers, we standardize the measures by using their ranks (instead of the original values) and align all the measures in the same direction. Specifically, for each QM, the facility performances are separated into 100 groups or "percentile ranks" ranging from 0.5 to 99.5 increasing by 1 where higher rank indicates a better score on a measure. To further differentiate facilities that performed exceptionally well or poorly, these percentile ranks (pRanks) were "normalized" or mapped from the uniform percentile rank distribution to a normal distribution (nRanks).

By using the transformation:

$$nRanks = \Phi^{-1}(pRanks \div 100) \times 19.4112 + 50\,,$$

the 0.5 and 99.5 percentile were first mapped to z-scores of the standard normal distribution. Scaling these 0 centered z-scores by a factor of 19.4112 and shifting by a value of 50, the normalized percentile ranks were centered at 50, with the lowest value achieving 0 and the highest 100.

Example: Suppose one of the QMs which measure the percentage of patients within a facility "passing" a threshold is right skewed (Figure 1). Using normal ranks allows many facilities to fall around the middle of the distribution, making extreme values more difficult to obtain. This method allows all measures to be scored in the same manner preventing different weighting on measures due to diverse distributions and scales. This method also manages to control outliers from having scores that differ extremely from the other facilities while recognizing that exceptionally high or low values should be distinguished.

Figure 1. Depiction of Normalization Algorithm

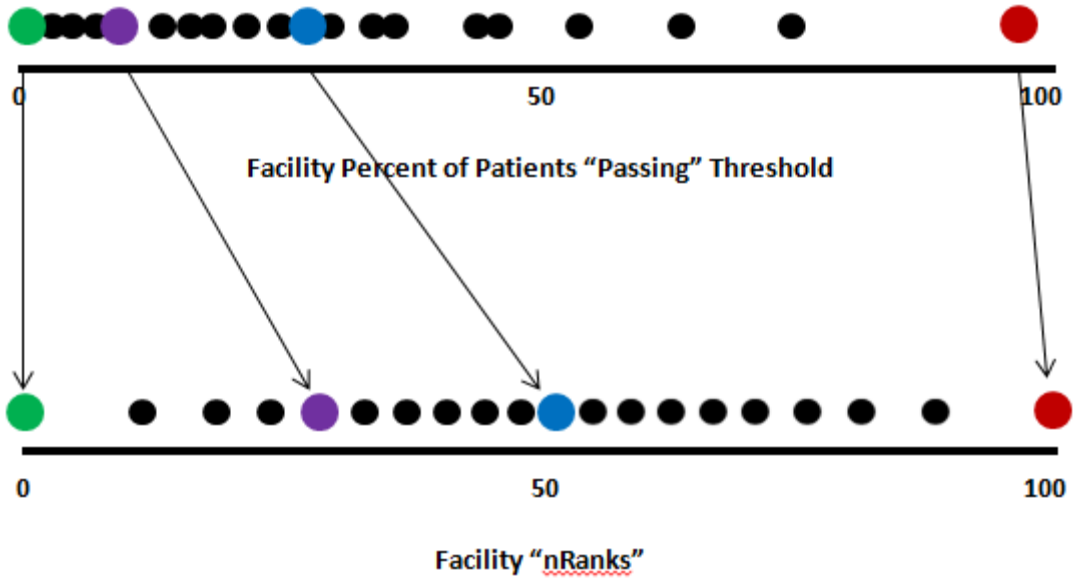


Table 1. Correlation of Normalized Measures

Measures	STrR	SHR	SMR	All Kt/V	Hypercal cemia	AVF	Catheter > 90
STrR	1.0000	***0.40139***	***0.21471***	0.08497	-0.00204	0.11354	0.15369
SHR		1.00000	***0.26229***	0.11016	0.00509	0.12759	0.18672
SMR			1.00000	0.07859	0.05328	0.16660	0.11062
All Kt/V				1.00000	***0.18577***	0.06416	0.13376
Hypercalcemia					1.00000	0.08786	0.04866
AVF						1.00000	***0.44751***
Catheter > 90							1.00000

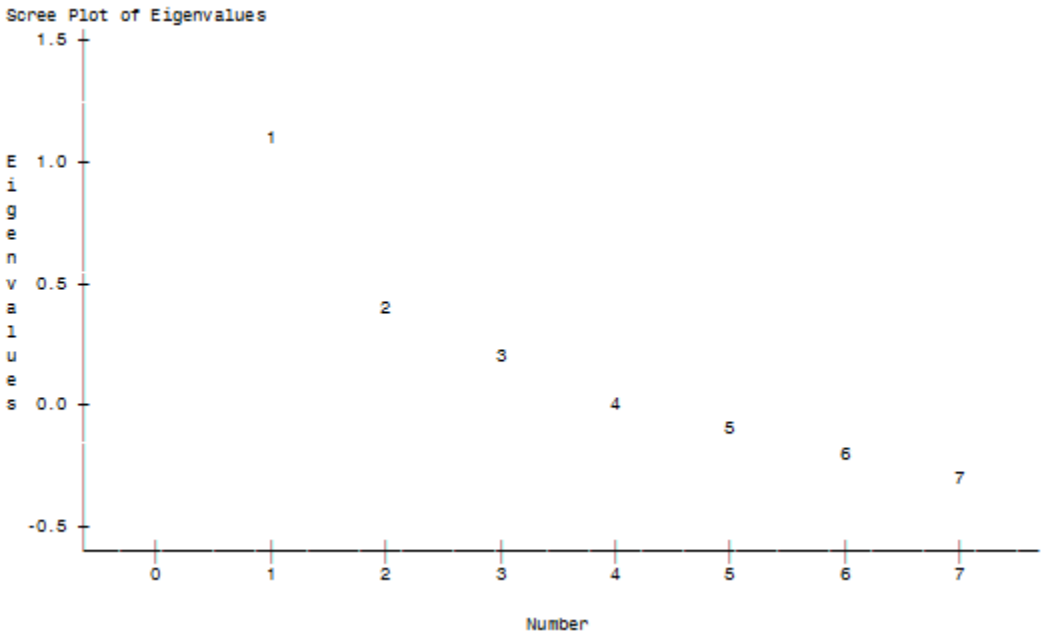
Based on January 2014 DFC data

Factor Analysis

When performing a factor analysis, we specify that our statistical software uses the method of principal components to extract the factors with loadings based on eigenvalue-eigenvector pairs of the sample covariance matrix. The resulting procedure is called the principal factor analysis, a common way of conducting factor analysis. We specify the prior communality estimates to substitute into the diagonal of the correlation matrix. The principal factor analysis uses squared multiple correlations (SMC) as priors. Finally, a rotation must be specified to obtain interpretable factor loadings (SAS/STAT 9.22 User's Guide).

The scree plot displayed in Figure 2 shows the eigenvalues associated with the correlation matrix of the measures in the December 2013 release dataset. One method of choosing the number of factors for data reduction is to take the factors before a breaking point in the plot (relatively large drop), and another, the positive eigenvalues (UCLA: Statistical Consulting Group). While there is a noticeable drop after the first eigenvalue (a global factor), a multiple factor solution allows the measurement of subgroups within the QMs. We observe a second, smaller break after the third eigenvalue which happens to be the cutoff between positive and negative eigenvalues. We investigate the three factor solution here for interpretable results.

Figure 2. Scree Plot of Eigenvalues



Both the orthogonal and oblique rotations were fit. The factor loadings from both methods were similar and yield the same interpretable results as to which QMs were associated with which domains. If

results had been different, the orthogonal rotation would have been the better method if the oblique solution had shown little correlation between factors. The QMs that are loaded highly on each of the three factors were allocated into 3 domains.

Quality Measure Domains

With the obtained factor loadings, the three respective empirically derived groups (domains) were also determined to correspond to related outcomes at the facility level. The three outcome measures for transfusions, mortality and hospitalization (STrR, SMR and SHR) formed the first grouping which was named the "Standardized Outcomes (SHR, SMR, STrR)". The arteriovenous fistula and catheter measures formed the second grouping which was named "Other Outcomes 1 (AV fistula, tunneled catheter)" The All Kt/V and hypercalcemia QMs formed the third grouping which was named "Other Outcomes 2 (Kt/V, hypercalcemia)". Together, these empirically derived groupings contain measures that are most correlated with one another, as indicated in the cells with the bolded correlation coefficients in Table 1. This is further evidence that grouped measures provide information on similar qualities about a facility.

Overall Star Rating for each Facility

To create the Star rating system, each domain is first given a score between 0 and 100 by averaging the normalized scores for measures within that domain. Facilities are given ratings as long as they have at least one measure in each domain. Facilities that served PD patients only (N=92 in the January 2014 data) do not have values for the two measures in the Other Outcomes 1 (AV fistula, tunneled catheter) Domain. These facilities were not excluded and instead were rated based on the average scores for the other domains. Among the 6,033 facilities in the January 2014 dataset, 542 (9% were unrated). In Table 2, the number and percentage of facilities with missing data is shown by the number of measures missing. Most facilities (81%) had all seven measures. Table 3 shows the number of facilities with missing data for each measure. The STrR measure was missing the most often in facilities.

Table 2. Number and Percent of Facilities Overall and Those Unrated by the Number of Measures Missing

# Measures Missing	# Facilities (%)	# Facilities Unrated (%)
0	4,903(81)	0 (0)
1	400 (7)	0 (0)
2	180 (3)	42 (23)
3	144 (2)	109 (76)
4	79 (1)	69 (87)
5	50 (1)	45 (90)
6	47 (1)	47 (100)
7	230 (4)	230 (100)
Total	6,033	542 (9)

Based on January 2014 DFC data

Table 3. Number and Percent of Facilities with Missing Data by Each Measure

Measures	# Facilities with Missing Data (%)
STrR	804 (13)
SHR	430 (7)
SMR	468 (8)
All Kt/V	386 (6)
Hypercalcemia	650 (11)
AVF	456 (8)
Catheter > 90 days	456 (8)

Based on January 2014 DFC data

After factor analysis is performed, missing values for facilities that qualified for ratings are assigned median pRanks and nRanks of 50. This method of imputation ensures that one measure is not too influential in the final rating. For instance, if one facility had an nRank of 100 for the catheter > 90 day measure and had no report of AVF, it would be unreasonable to assume that the Other Outcomes 1 (AV fistula, tunneled catheter) Domain should be given an average score of 100. By imputing 50 (the average) for the AVF measure, we instead give the domain a score of 75, still well above average, but conservative enough to limit catheter > 90 days measure from being too influential.

A final score between 0 and 100 is then created by averaging the three domain scores.

Finally, to recognize high and low performances, facilities receive stars in the following way:

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

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

In the January 2014 release dataset, we observed a noticeable systematic improvement of all average measure values with higher star rating (Table 4).

Table 4. Average Measure Values Within Overall Star Rating

Measure	Average Measure Values Within Star Rating				
	★	★★	★★★	★★★★	★★★★★
STrR	1.50	1.20	1.00	0.81	0.63
SHR	1.28	1.12	0.99	0.86	0.75
SMR	1.34	1.11	1.02	0.93	0.84
All Kt/V	75.5	81.8	86.8	89.5	92.3
Hypercalcemia	5.7	4.6	3.4	2.3	1.8
AVF	48.6	56.0	62.1	67.3	73.2
Catheter > 90	20.3	14.7	10.6	7.6	5.2

Based on January 2014 DFC data

Conclusions

This methodology report presents an overview of the DFC Star Rating of facilities based on the groupings of correlated quality measures that are currently reported on the Medicare DFC website. In future years, when reported DFC measures change, the general algorithm described here will be used to update measure domains used to produce the rating. For the implementation of the Star System with January 2014 data, average measure values are consistently better with higher Overall Star Rating (Table 4). The analysis of ratings over time was limited because data for some measures have only been available recently. However, the data available showed evidence that the ratings would not behave erratically over time. An advantage to the Star Rating, is the grouping of QMs based on systematic empirical methods, specifically, factor analysis. This method limits the possibility of overweighting QMs that measure similar qualities of facility care. Finally, the Star Rating is updated annually, to align with the annual updates of the standardized measures..

References

Johnson, R. A. & Wichern, D. W. (2007). Chapter 9: Factor Analysis and Inference for Structured Covariance Matrices (6th ed). *Applied Multivariate Statistical Analysis*. Pearson.

SAS/STAT(R) 9.22 User's Guide. Retrieved April 26, 2014, from http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_factor_sect029.htm

UCLA: Statistical Consulting Group. Factor Analysis Using SAS PROC FACTOR. Retrieved April 26, 2014, from http://www.ats.ucla.edu/stat/sas/notes2/