

# DRAFT

## *Anemia of chronic kidney disease: Dialysis facility standardized transfusion ratio*

### 3b Measure Justification

#### Importance

- ◆ **High Impact Aspect of Health Care**
  - **Demonstrated high impact aspect**

1a1.1 Select from the following all that apply:

    - *Affects large numbers*
    - *A leading cause of morbidity/mortality*
    - *Frequently performed procedure*
    - **High resource use: YES**
    - **Patient/societal consequences of poor quality: YES**

- **Summary of evidence of high impact**

1a3. Provide epidemiological or resource use data

Safety concerns arising from clinical trials of ESA treatment of anemia of chronic kidney disease (CKD) have led to recent changes in FDA recommendations on ESA use in patients with CKD. In addition, changes in financial incentives for treatment of anemia following the implementation of the revised Medicare ESRD Prospective Payment System have further heightened concerns in the dialysis community that patients with CKD-related anemia may be denied adequate access to ESAs for prevention of red blood cell transfusion. This concern has been further amplified by recently reported trends in anemia management in US chronic dialysis patients, demonstrating rapid declines in achieved hemoglobin from mid-2010 to the present.

The risks associated with aggressive treatment of anemia of CKD with ESAs have been well documented in KDIGO Anemia Management Guidelines as well as in updated FDA package insert information for ESAs. In contrast, the effect of anemia management paradigms that target to lower hemoglobin levels, and generally use less ESA, on transfusion risk is less well defined. Several clinical interventional trials comparing higher vs. lower hemoglobin targets have shown higher transfusion rates in those patients randomized to lower hemoglobin targets. The importance of these observations is limited by lack of predefined criteria for use of blood transfusion in most studies.

It has been postulated that a national trend toward increased use of transfusions in dialysis patients would adversely affect the supply of blood available for acute injuries and surgical procedures. Lastly, greater exposure to human leukocyte antigens, present in transfused blood, may increase anti-HLA antibodies in kidney transplant candidates, resulting in reduced access to kidney transplantation.

The inverse relationship between achieved hemoglobin and transfusion events has been reported previously for Medicare dialysis patients (Ma, J Am Soc Nephrol, 1999) and for non-dialysis CKD patients treated in the Veterans Administration system (Lawler, Clin J Am Soc Nephrol, 2010)

Unpublished analyses of Medicare Claims data presented at CMS Technical Expert Panel in May 2012 demonstrate an inverse association between achieved hemoglobin and subsequent transfusion risk using more recent data from 2008-2011.

In early 2012, a highly publicized USRDS study presented at the NKF Clinical meeting reported increased dialysis patient transfusion rates in 2011 compared to 2010.

UM-KECC and Arbor Research collaborators have recently presented an analysis of transfusion events in Medicare dialysis patients from 2009-2011, observing increased transfusions in 2011, although the magnitude of change in transfusion rates was much lower than reported by the USRDS.

o **Citations**

*1a.4. Provide citations for the evidence described above*

- Hollenbeak et. al. The Impact of End-Stage Renal Disease Transfusion Demand on Blood Utilization and Blood Supply in the United States Health Outcomes Research in Medicine Volume 3, Issue 2, May 2012, Pages e67–e77
- Liu et. al. Development of a Facility-Level Transfusion Quality of Care Metric, 2012 American Society of Nephrology Annual Kidney Week
- Ibrahim HN, et. al. Temporal Trends in red blood transfusion among US dialysis patients, 1992-2005. Am J Kidney Dis 2008; 52: 1115
- U.S. Renal Data System, USRDS 2012 Annual Data Report: Atlas of End-Stage Renal Disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2012.
- FDA Drug Safety Communication: Modified dosing recommendations to improve the safe use of Erythropoiesis-Stimulating Agents (ESAs) in chronic kidney disease.  
<http://www.fda.gov/Drugs/DrugSafety/ucm259639.htm>
- Highlights of prescribing information: Epogen (epoetin alfa)  
[http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2011/103234Orig1s5166\\_103234Orig1s5266lbl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/103234Orig1s5166_103234Orig1s5266lbl.pdf)
- Highlights of prescribing information: Aranesp (darbepoetin alfa)  
[http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2011/103951Orig1s5173\\_103951Orig1s5258lbl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/103951Orig1s5173_103951Orig1s5258lbl.pdf)
- Kidney Disease: Improving Global Outcomes (KDIGO) Anemia Work Group. KDIGO Clinical Practice Guideline for Anemia in Chronic Kidney Disease. Kidney inter., Suppl. 2012; 2: 279–335.
- Besarab A, Bolton WK, Browne JK, et al. "The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin." The New England journal of medicine (1998) 339:584-90. PMID: 9718377
- Drüeke TB, Locatelli F, Clyne N, et al. "Normalization of hemoglobin level in patients with chronic kidney disease and anemia." The New England journal of medicine (2006) 355:2071-84. PMID: 17108342
- Foley RN, Curtis BM, Parfrey PS. "Hemoglobin targets and blood transfusions in hemodialysis patients without symptomatic cardiac disease receiving erythropoietin therapy." Clinical journal of the American Society of Nephrology : CJASN (2008) 3:1669-75. PMID: 18922988
- Lawler EV, Bradbury BD, Fonda JR, et al. "Transfusion burden among patients with chronic kidney disease and anemia." Clinical journal of the American Society of Nephrology : CJASN (2010) 5:667-72. PMID: 20299366
- Ma JZ, Ebben J, Xia H, et al. "Hematocrit level and associated mortality in hemodialysis patients." Journal of the American Society of Nephrology : JASN (1999) 10:610-9. PMID: 10073612

- Pfeffer MA, Burdmann EA, Chen CY, et al. "A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease." The New England journal of medicine (2009) 361:2019-32. PMID: 19880844
- Association between recombinant human erythropoietin and quality of life and exercise capacity of patients receiving haemodialysis. Canadian Erythropoietin Study Group." BMJ (Clinical research ed.) (1990) 300:573-8. PMID: 2108751
- Hirth R, Turenne M, Wheeler JRC, Nahra T, Sleeman K, Zhang W, Messana JM. Did the Dialysis Prospective Payment System Result in more patients receiving transfusions? Abstract presented at ASN Renal Week in San Diego, November 2012.
- Ibrahim HN, Skeans MA, Li Q, Ishani A, Snyder JJ. Blood transfusions in kidney transplant candidates are common and associated with adverse outcomes. Clin Transplant (2011):25;653-659

◆ **Opportunity for Improvement**

- **Briefly explain the benefits envisioned by use of this measure**

*1b.1. (Quality improvement anticipated)*

The proposed standardized transfusion ratio (STrR) measure would be used to monitor relative transfusion rates among dialysis facilities, identifying facilities that may be unnecessarily using blood transfusions to treat anemia. Implementation of this measure should contribute to the preservation of limited blood supply resources. In addition, avoidance of unnecessary blood transfusion will help protect patient access to kidney transplantation, by reducing the exposure of kidney transplant candidates to human tissue antigens, thereby reducing the risk for immune sensitization to these antigens.

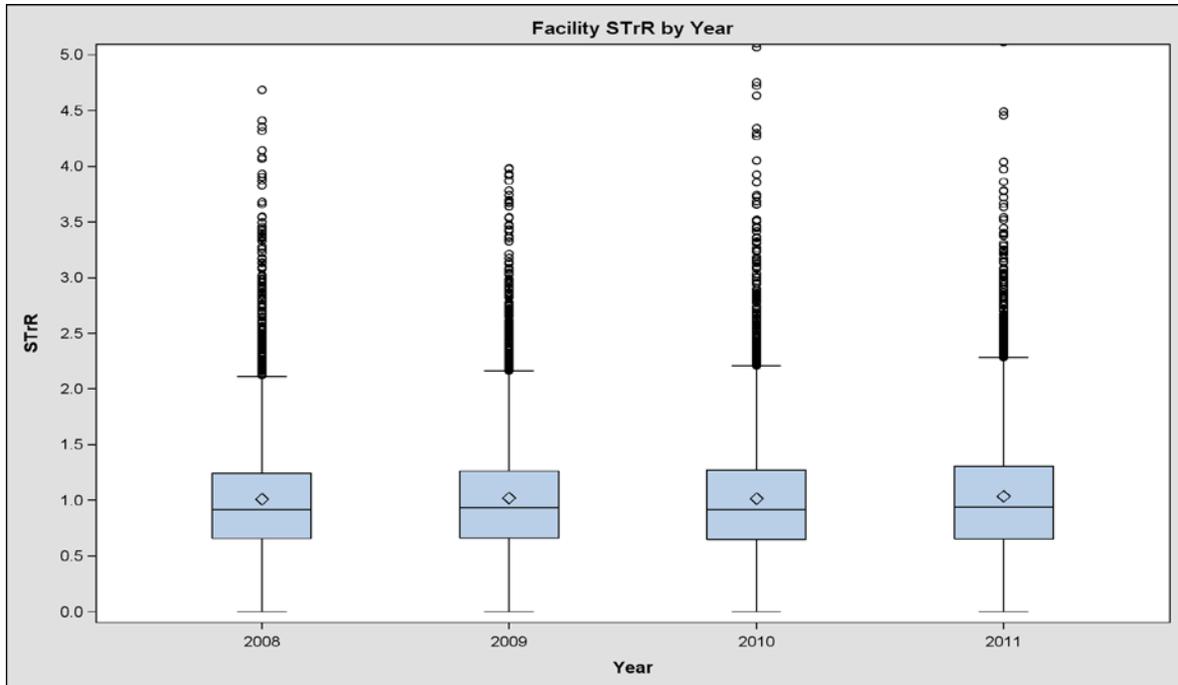
- **Summary of data demonstrating performance gap**

*1b.2. (Variation or overall less than optimal performance across providers)*

The STrR is a facility-level measure, comparing the observed number of red blood cell transfusions counts at a facility with the number of transfusions that would be expected under a national norm, after accounting for the patient characteristics within each facility. Standardized transfusion ratios vary across facilities. The following table shows the distribution of STrR using Medicare claims data for 2008-2011. As implemented in Standardized Hospitalization Ratio measure (NQF #1463 <http://www.qualityforum.org/QPS/1463>) facilities with less than 5 patient years at risk are excluded from this analysis.

Year	# of facilities	Mean STrR	Standard Error	Facility percentile				
				10th	25th	50th	75th	90th
2008	4802	1.010	0.554	0.438	0.660	0.922	1.243	1.666
2009	5077	1.023	0.545	0.456	0.665	0.936	1.264	1.685
2010	5261	1.017	0.559	0.437	0.646	0.917	1.273	1.695
2011	5407	1.035	0.566	0.446	0.653	0.942	1.306	1.718

Graphically, the distribution is shown with the following box plot. For better understanding of the distribution, the vertical axis is truncated at 5.



o **Citations**

1b.3. Provide citations for the evidence described above

- Unpublished analysis on draft STR measure based on Medicare claims done by Arbor Research Collaborative for Health and Kidney Epidemiology and Cost Center- University of Michigan.

o **Summary of data on disparities by population group**

1b.4. Summarize evidence found that demonstrates any disparities. Describe groups in which disparities exist.

Analyses of the STR by race, sex and ethnicity indicate relatively little variation and no substantial disparities among these groups. Although females are somewhat more likely to receive transfusions than males, analyses showed that a model with race and sex included and a model without these variables yielded very similar results for the facility STR measure as well as for the parameter estimates for other variables. The table below shows the parameter estimates for the race, sex and ethnicity variables based on a model that included these variables along with other covariates.

Model with sex, race, ethnicity included along with other covariates			
Parameter	Estimate	Standard Error	P value
Females	0.08126	0.00672	<.0001
Native American*	-0.15707	0.01795	<.0001

Model with sex, race, ethnicity included along with other covariates			
Parameter	Estimate	Standard Error	P value
Asian*	-0.23275	0.01065	<.0001
Black*	-0.0816	0.00464	<.0001
Other Race*	-0.05411	0.01843	0.0033
Hispanic ¥	-0.1919	0.00662	<.0001

\*White used as reference

¥Non-Hispanic used as reference

o **Citations**

1b.5. Provide citations for the evidence described above

Unpublished analysis on draft STrR measure based on Medicare claims done by Arbor Research Collaborative for Health and Kidney Epidemiology and Cost Center- University of Michigan.

**Evidence to Support Measure Focus**

o **Structure-process-outcome relationship**

1c.1. Briefly state the measure focus (for example, health outcome, intermediate clinical outcome, process, structure) Then, identify the appropriate links (for example, structure-process-health outcome, process-health outcome, intermediate clinical outcome-health outcome)

The Standardized Transfusion Ratio represents an outcome measure.

In the ESRD population, blood transfusion has been linked to survival indirectly via patient access to transplantation. Studies have shown superior patient survival with kidney transplantation compared to chronic dialysis (Wolfe, et al NEJM).

Blood transfusion has been shown to increase anti-HLA antibodies in chronic dialysis patients, decreasing access to kidney transplantation. (Chapter 4, KDIGO Anemia Management Guidelines) Furthermore, Ibrahim, et al studied 43,025 patients added to the kidney transplant waitlist from 1999-2004, using USRDS data. They evaluated the impact of receiving one or more blood transfusion after kidney transplant listing on panel reactive antibody% (PRA). Over the years 1999-2004, 26-30% of patients listed for kidney transplant received one or more blood transfusion after listing. Ibrahim, et al calculated the one year and three year cumulative incidence of transfusions while on the waiting list at 10.8% and 27.7% respectively. Receiving pre-transplant transfusion was associated with higher odds of PRA% elevation. For men, post-listing transfusion was associated with an odds ratio of 1.77 and 1.67 for having a PRA  $\geq$  20% and  $\geq$  80% at time of transplantation, respectively. For parous women, odds ratios were 1.62 and 1.89 for PRA  $\geq$  20% and  $\geq$  80% at time of transplantation, respectively.

In addition, unnecessary use of blood products in this population will likely have a negative impact on the health outcomes of other patient populations by reducing a rate-limiting health resource needed for treatment of other life-threatening conditions.

KDIGO Anemia Guidelines 2012: Guideline 4.1.1: When managing chronic anemia, we recommend avoiding, when possible, red cell transfusions to minimize the general risks related to their use. (1B)

Reference:

Wolfe, Robert, Ashby, Valarie, Milford, Edgar et al. Comparison of Mortality in all Patients on Dialysis, Patients on Dialysis Awaiting Transplantation, and Recipients of a First Cadaveric Transplant. The New England Journal of Medicine (1999) 341:1725-30.

Ibrahim HN, Skeans MA, Li Q, Ishani A, Snyder JJ. Blood transfusions in kidney transplant candidates are common and associated with adverse outcomes. Clin Transplant (2011) 25;653-659.

o **Type of evidence**

- 1c.2. Describe the type of evidence, selecting from the following list all that apply:
- **Clinical practice guideline: YES**
- **Selected individual studies (rather than entire body of evidence): YES**
- **Systematic review of body of evidence (other than within guideline development):**
- **Other (state type of evidence)**

o **Directness of evidence to the specified measure**

1c.4. State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population.

There are no differences between the body of evidence and the measure focus in topic, population, and outcomes.

o **Quantity of studies in the body of evidence**

1c.5. Total number of studies, not articles

The body of evidence is summarized in Chapter 4, KDIGO Guidelines (references 190-229 in the KDIGO publication) and in the two individual studies cited here.

o **Quality of body of evidence**

1c.6. Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address:

- a) Study design/flaws
- b) Directness/indirectness of the evidence to this measure (for example, interventions, comparisons, outcomes assessed, population included in the evidence)  
*Imprecision/wide confidence intervals due to few patients or events)*

The relevant KDIGO Guideline was given a “moderate” grade for quality of evidence.

o **Consistency of results across studies**

1c7. Summarize the consistency of the magnitude and direction of the effect across studies

The majority of the reviewed studies agreed transfusions should be avoided when possible.

- **Net benefit**

*1c8. Provide estimates of effect for benefit/outcome, identify harms addressed and estimates of effect, and net benefit---benefit over harms across studies. Please include results of business/social/economic case for the measure.*

N/A: the STrR is an outcome measure.

- **Grading of strength/quality of the body of evidence**

*1c9, 1c10, 1c11, 1c13, 1c14. Please address:*

- *Indicate if the body of evidence has been graded*
- *If the body of evidence was graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias*
- *System used for grading the body of evidence*
- *Grade assigned to the body of evidence*  
*Summary of controversy/contradictory evidence*

N/A: the STrR is an outcome measure.

- **Citation**

*1c15. Provide citations for the evidence described above*

See 1a.4 above for citations.

- **Guideline recommendation**

*1c16. Quote verbatim, the specific guideline recommendation (Including guideline number and/or page number)*

- **KDIGO Anemia Guidelines 2012: Guideline 3.2:** In initiating and maintaining ESA therapy, we recommend balancing the potential benefits of reducing blood transfusions and anemia-related symptoms against the risks of harm in individual patients (e.g., stroke, vascular access loss, hypertension). (1B).
- **KDIGO Anemia Guidelines 2012: Guideline 4.1.1:** When managing chronic anemia, we recommend avoiding, when possible, red cell transfusions to minimize the general risks related to their use. (1B)
- **KDIGO Anemia Guidelines 2012: Guideline 4.1.3:** When managing chronic anemia, we suggest that the benefits of red cell transfusions may outweigh the risks in patients in whom (2C):
  - **ESA therapy is ineffective** (e.g., hemoglobinopathies, bone marrow failure, ESA resistance)
  - **The risks of ESA therapy may outweigh its benefits** (e.g., previous or current malignancy, previous stroke)

- **Citation**

*1c17. Provide citations for the clinical practice guideline quoted above*

Kidney Disease: Improving Global Outcomes (KDIGO) Anemia Work Group. KDIGO Clinical Practice Guideline for Anemia in Chronic Kidney Disease. *Kidney inter., Suppl.* 2012; 2: 279–335.

- **URL**

*1c18. National Guideline Clearinghouse or other URL*

[http://www.kdigo.org/clinical\\_practice\\_guidelines/pdf/KDIGO-Anemia%20GL.pdf](http://www.kdigo.org/clinical_practice_guidelines/pdf/KDIGO-Anemia%20GL.pdf)

o **Grading of strength of recommendation**

1c191 1c21, 1c23. Please address:

- *Has the recommendation been graded?*
- *System used for grading the strength of guideline recommendation (USPSTF, GRADE, etc.) Grade assigned to the recommendation*

*Grading system (1- We recommend, 2- We suggest) combined with a 4 category quality of evidence grading (A, B, C, D).*

The KDIGO Guidelines used the GRADE system; the grades given are listed above with the relevant guidelines. The definitions used by KDIGO are listed below.

**\_\_\_ NOMENCLATURE AND DESCRIPTION FOR RATING GUIDELINE RECOMMENDATIONS**

Within each recommendation, the strength of recommendation is indicated as **Level 1**, **Level 2**, or **Not Graded**, and the quality of the supporting evidence is shown as **A**, **B**, **C**, or **D**.

Grade*	Implications		
	Patients	Clinicians	Policy
Level 1 'We recommend'	Most people in your situation would want the recommended course of action and only a small proportion would not.	Most patients should receive the recommended course of action.	The recommendation can be evaluated as a candidate for developing a policy or a performance measure.
Level 2 'We suggest'	The majority of people in your situation would want the recommended course of action, but many would not.	Different choices will be appropriate for different patients. Each patient needs help to arrive at a management decision consistent with her or his values and preferences.	The recommendation is likely to require substantial debate and involvement of stakeholders before policy can be determined.

\*The additional category 'Not Graded' was used, typically, to provide guidance based on common sense or where the topic does not allow adequate application of evidence. The most common examples include recommendations regarding monitoring intervals, counseling, and referral to other clinical specialists. The ungraded recommendations are generally written as simple declarative statements, but are not meant to be interpreted as being stronger recommendations than Level 1 or 2 recommendations.

Grade	Quality of evidence	Meaning
A	High	We are confident that the true effect lies close to that of the estimate of the effect.
B	Moderate	The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
C	Low	The true effect may be substantially different from the estimate of the effect.
D	Very Low	The estimate of effect is very uncertain, and often will be far from the truth.

o **Rationale for using this guideline over others**

1c24. *If multiple guidelines exist, describe why the guideline cited was chosen. Factors may include rigor of guideline development, widespread acceptance and use, etc.*

N/A: The KDIGO Guidelines are the most recent and relevant for the dialysis population.

o **Overall assessment of the body of evidence**

1c25, 1c26, 1c.27. *Based on the NQF descriptions for rating the evidence, what was your assessment of the following attributes of the body of evidence?*

- **Quantity: High**
- **Quality: Moderate**
- **Consistency: High**

## Reliability and Validity – Scientific Acceptability of Measure Properties

### ◆ Reliability Testing

#### ○ Data sample

2a2.1. Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.

The reliability of the Standardized Transfusion Ratio was assessed using data on transfusions among ESRD patients over a four year period of 2008-2011. The table below shows the number of facilities, patients, total count of transfusions and total patient years at risk for each year. Also, we calculate unadjusted or raw transfusion rate per year defined as total transfusions divided by total patient years at risk.

Year	# facilities	# of Patients	Total transfusions	Total Patients Years at risk	Raw Transfusion Rate per 100 patient years at risk*
2008	5178	351217	89476	205325.8	43.58
2009	5468	368348	92624	213126.0	43.46
2010	5636	387077	93409	222069.8	42.06
2011	5757	397621	102374	225890.8	45.32

\*This analysis includes all facilities for the given year.

Data for the measure are derived from an extensive national ESRD patient database, which is derived from Program Medical Management and Information System (PMMIS/REMIS), Medicare claims, the Standard Information Management System (SIMS) database maintained by the 18 ESRD Networks, the CMS Annual Facility Survey (Form CMS-2744), Medicare dialysis and hospital payment records, the CMS Medical Evidence Form (Form CMS-2728), transplant data from the Organ Procurement and Transplant Network (OPTN), the Death Notification Form (Form CMS-2746), the Nursing Home Minimum Dataset, and the Social Security Death Master File. The database is comprehensive for Medicare patients. Information on transfusions is obtained from Medicare Inpatient and Outpatient Claims Standard Analysis Files (SAFs).

2a2.2.

#### ○ Analytic methods

2b2.2. Describe method of validity testing and rationale; if face validity, describe systematic assessment

To assess reliability, we assessed the degree to which the measures were consistent year to year. If one looks at two adjacent time intervals, one should expect that a reliable measure will exhibit correlation over these periods since large changes in patterns affecting the measure should not occur for most centers over shorter periods. Year to year variability in the measure values was assessed across the years 2008, 2009, 2010 and 2011 based on dialysis centers for which a 2012 Dialysis Facility Report (DFR) is available.

- **Testing results**

*2a2.3. Provide reliability statistics and assessment of adequacy in the context of norms for the test conducted*

The correlation between the measure across adjacent years (2008 vs. 2009, 2009 vs. 2010, and 2010 vs. 2011) are 0.41, 0.40 and 0.42 respectively, indicating a tendency for facilities with higher or lower transfusion rates in one year to have higher or lower transfusion rates in the following year. These correlations were highly significant. The measure is based on complete data and is not subject to judgment or rater variability. Hence the measures of inter-rater variability are not relevant here.

- ◆ **Validity Testing**

- **Data sample**

*2b2.1. Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*

We developed STrR measure using claims data for 2008-2011 for dialysis patients. Refer to section 2a2.1 for the detailed data description.

- **Analytic method**

*2b2.2. Describe method of validity testing and rationale; if face validity, describe systematic assessment*

We examined this measure's correlations with the other measures of quality among ESRD population and reported significant correlation estimates. We assessed the validity of the measure through various comparisons of this measure with other quality measures in use, and in May 2012 there was an assessment of face validity based on polling of a CMS Technical Expert Panel (TEP).

- **Testing results**

*2b2.3. (Provide statistical results and assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*

6/6 voting members of CMS' Technical Expert Panel voted to recommend development of a facility-level Standardized Transfusion Ratio measure. The consensus recommendation of that clinical expert panel included the recommendation to include risk adjustment for conditions that are associated with an increased risk of blood transfusion such as hereditary anemia, chronic bone marrow failure conditions and active cancer.

The validity of the STrR measure is supported by its association with other known quality measures, which include both dialysis facility outcomes and practices. For year 2011, we find that the measure is positively correlated with two health outcome measures: the one-year Standardized Hospitalization Ratio for Admissions ( $r = .42, p < .0001$ ) and the one-year Standardized Mortality Ratio ( $r = .25, p < .0001$ ). That is, facilities with more transfusions than would be expected based on national rates also have higher mortality and more hospital admissions than would be expected based on national rates.

We also checked the correlation with average hemoglobin value of all ESA-treated dialysis patients and ( $r = -.23, p < .0001$ ) a negative correlation indicates that lower values of hemoglobin are associated with higher values of STrR.

Similarly, a positive correlation with the percent of patients with Hgb < 10 ( $r = .22, p < .0001$ ) indicates that higher % of patients with Hgb < 10 is associated with higher STRR.

Furthermore, the STRR is correlated with catheter use ( $r = .18, p < .0001$ ), indicating that higher values of STRR are associated with increased use of catheters. The STRR is negatively correlated with the percentage of patients having a Urea Reduction Ratio (URR) of at least 65% ( $r = -.13, p = .0003$ ) and using a fistula ( $r = -.11, p < .0001$ ). That is, higher values of STRR are associated with lower rates of URR and fistula use.

## References

- 2012 Anemia Management TEP Summary Report  
<http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TechnicalExpertPanels.html>

## Exclusions

### Data sample for analysis of exclusions

*2b3.1. Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*

Comorbidity exclusions of the measure were assessed using data on ESRD patients for 2011. For 2011, the data represents dialysis patients at 5,757 facilities and a total count of 102,374 transfusions. Refer to section 2a2.1 for the detailed data description.

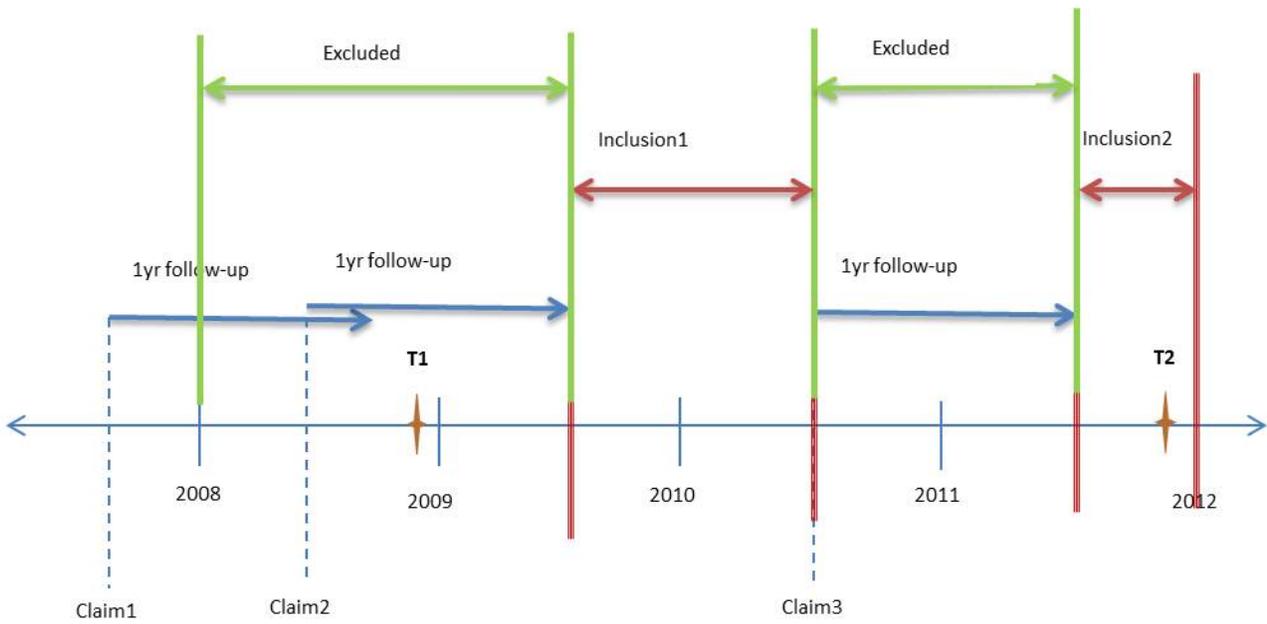
#### o Analytic method

*2b3.2. Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference*

Patient-months are excluded from the measure if on the first of the month the patient is fewer than 90 days since first ESRD service date due to incompleteness of data and differing ESA dose practices. In addition, patients that are less than 18 years of age are excluded due to the relatively small number of pediatric patients treated at most facilities. Also, all transfusions associated with transplant hospitalization are excluded.

Patients are also excluded if they have a Medicare claim for hemolytic and aplastic anemia, solid organ cancer (breast, prostate, lung, digestive tract and others), lymphoma, carcinoma in situ, coagulation disorders, multiple myeloma, myelodysplastic syndrome and myelofibrosis, leukemia, head and neck cancer, other cancers (connective tissue, skin, and others), metastatic cancer, sickle cell anemia within one year of their patient at risk time. Since these comorbidities are associated with higher risk of transfusion and require different anemia management practices that this measure is not intended to address, every patient's risk window is modified to have at least 1 year of claim free period. We assessed the predictive power of comorbidities on future transfusions, as a function of the time interval between development of the comorbidity and the occurrence of the transfusion by performing multivariate logistic regression with transfusion count as the dependent variable. Results showed that 1-year look back period for each of the above mentioned comorbidities was the most predictive of one or more RBC transfusions.

The following figure describes the inclusion and exclusion period of a hypothetical patient.



In the figure above, a hypothetical patient has patient years at risk at a facility from 1/1/2008 to 12/31/2011. Review of Medicare claims identified presence of one or more exclusion comorbidities (see above and Appendixe) in 2007 (Claim1), 2008 (Claim2) and 2010 (Claim3). Each claim is followed by a one year exclusion period. The revised inclusion periods are defined as risk windows with at least 1 year of claim-free period ( Inclusion1 and Inclusion2 in Figure1). The patient has two transfusion events, marked as T1 and T2 in late 2008 and late 2011 respectively. However, since T1 falls in the exclusion period, it will not be counted towards the facility's transfusion count as presence of exclusion comorbidity claims within a year might have increased the risk of transfusion unrelated to dialysis facility anemia management practice. However, T2, which occurs in late 2011 and in Inclusion2 period, will be counted since there is at least a year gap between this transfusion event and the last claim observed

- **Results**

*2b3.3. Provide statistical results for analysis of exclusions (for example, frequency, variability, sensitivity analyses)*

Multivariate logistic regression with transfusion count as the dependent variable was performed to assess the predictive power of comorbidities on future transfusions, as a function of the time interval between development of the comorbidity and the occurrence of the transfusion. Transfusion count was coded as a binary variable (1 if transfusion). Result using 2011 data showed that 1-year look back period for each of the above mentioned comorbidities was a significant predictor of RBC transfusion events with odds ratio ranging from 1.2 to 3.2.

- ◆ **Risk Adjustment Strategy**

- **Data/ sample**

*2b4.1. Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Delete row if measure is not risk adjusted.*

Risk adjustment for this measure is based on a Cox model using data on transfusions among ESRD patients over a four year period of 2008-2011 national data. Refer to section 2a2.1 for the detailed data description.

- **Analytic method**

*2b4.2. Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables.*

The denominator of the "STrR" uses expected transfusions calculated from a Cox model (Cox, 1972) as extended to handle repeated events (Lawless and Nadeau, 1995; Lin et al., 2000; Kalbfleisch and Prentice, 2002). For computational purposes, we adopt a model with piecewise constant baseline rates (e.g. Cook and Lawless, 2007) and computational methodology as developed in Liu, Schaubel and Kalbfleisch (2010). A stage 1 model is first fitted to the national data with piecewise-constant baseline rates stratified by facility; transfusion rates are adjusted for patient age, diabetes, duration of ESRD, nursing home status, BMI at incidence, comorbidity index at incidence, and calendar year. This model allows the baseline transfusion rates to vary between strata (facilities), but assumes that the regression coefficients are the same across all strata; this approach is robust to possible differences between facilities in the patient mix being treated. The linear predictor for each patient based on the regression coefficients in the stage 1 model is used to compute a risk adjustment factor that is then used as an offset in the stage 2 model.

The patient characteristics included in the stage 1 model as covariates are age (18-24 years old, 25-44 years old, 45-59 years old, 60-74 years old, or 75+ years old), cause of ESRD (diabetes or other), nursing home status, BMI at

incidence, comorbidity index at incidence, duration of ESRD (91 days-6 months, 6 months-1 year, 1-2 years, 2-3 years, 3-5 years, or 5+ years as of the period start date) and calendar year. Nursing home status is identified as in or not in a nursing home in the previous calendar year. The comorbidity index is calculated as a weighted linear combination of comorbidities reported on the Medical Evidence Form (CMS-2728) namely alcohol dependence, atherosclerotic heart disease, cerebrovascular disease, chronic obstructive pulmonary disease, congestive heart failure, diabetes, diabetes (currently on insulin), drug dependence, inability to ambulate, inability to transfer, malignant neoplasm, cancer, other cardiac disease, peripheral vascular disease, tobacco use (current smoker) using the same weights as used for Standardized Hospitalization Ratio (<http://www.dialysisreports.org/pdf/esrd/public/shrmodel.xls>; NQF #1463 <http://www.qualityforum.org/QPS/1463>). BMI is included as a log-linear term. Categorical indicator variables are included as covariates in the stage 1 model to flag records with missing values for cause of ESRD, comorbidity index, and BMI. These variables have a value of 1 if the patient is missing the corresponding piece of information and a value of 0 otherwise. Another categorical indicator variable included as a covariate to flag records where the comorbidity index is 0 has a value of 1 if the patient has a comorbidity index of 0 and a value of 0 otherwise. Beside main effects, some two way interaction terms are also included in the model based on their clinical and statistical significance.

o **Testing results**

*2b4.3. Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata. Delete row if measure is not risk adjusted.*

The parameter estimates as well as the corresponding standard error and a p-value indicating if the coefficient is significantly different from 0, resulting from the Cox Model are shown below.. All covariates have face validity from a clinical perspective and are based on the list of covariates used in Standardized Hospitalization Ratio (NQF #1463 <http://www.qualityforum.org/QPS/1463>). With the exception of Cause of ESRD missing, all main effects are statistically significant at 0.05 level.

**Table 1: Parameter Estimates of Transfusion Events for Medicare-Covered Dialysis Patients**

Parameter	Level	Type	Estimate	Standard Error	p value
Age		Categorical (60-74 is ref)			
15-24 years old		Categorical (60-74 is ref)	0.0045	0.0187	0.8099
25-44 years old		Categorical (60-74 is ref)	-0.2295	0.0077	<.0001
45-59 years old		Categorical (60-74 is ref)	-0.1547	0.0063	<.0001
75 or older		Categorical (60-74 is ref)	-0.0122	0.0063	0.0539
Diabetes		Categorical (0 versus 1)	-0.0633	0.0143	<.0001
Cause of ESRD Missing		Categorical (0 versus 1)	-0.0164	0.0212	0.439
Patient in Nursing Home		Categorical (0 versus 1)	0.5788	0.0053	<.0001
Log of BMI		Continuous	-0.1859	0.0069	<.0001
BMI Missing		Categorical (0 versus 1)	0.1073	0.0097	<.0001
Comorbidity Index		Continuous	0.3624	0.0075	<.0001

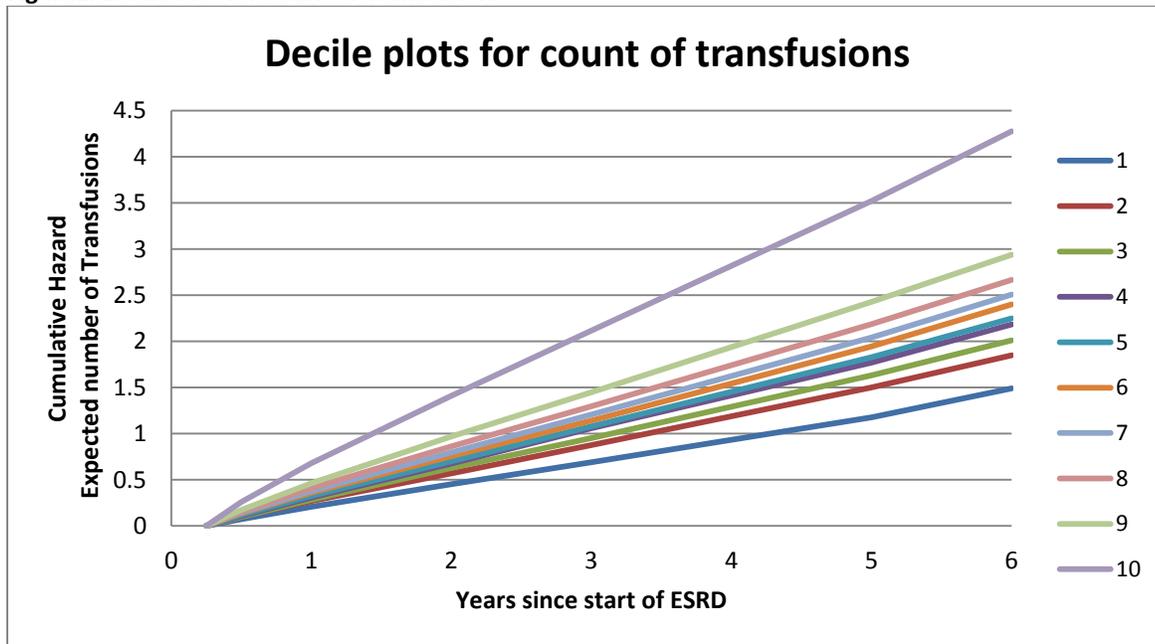
Parameter	Level	Type	Estimate	Standard Error	p value
Comorbidity Index of 0		Categorical (0 versus 1)	-0.1280	0.0057	<.0001
Comorbidity Index Missing		Categorical (0 versus 1)	-0.1052	0.0274	0.0001
Year	2009	Categorical	-0.0196	0.0048	<.0001
Year	2010	Categorical	-0.0512	0.0048	<.0001
Year	2011	Categorical	0.0169	0.0048	0.0004
Duration of ESRD*Diabetes	6 months-1 year	Interaction (Duration of ESRD in Diabetes)	0.0531	0.0173	0.0021
Duration of ESRD*Diabetes	1-2 years	Interaction (Duration of ESRD in Diabetes)	0.0794	0.0159	<.0001
Duration of ESRD*Diabetes	2-3 years	Interaction (Duration of ESRD in Diabetes)	0.0966	0.0163	<.0001
Duration of ESRD*Diabetes	3-5 years	Interaction (Duration of ESRD in Diabetes)	0.0526	0.0155	0.0007
Duration of ESRD*Diabetes	5+ years	Interaction (Duration of ESRD in Diabetes)	0.0322	0.0150	0.0322
Age*Diabetes	15-24 years old	Interaction (Age in Diabetes)	0.3414	0.0824	<.0001
Age*Diabetes	25-44 years old	Interaction (Age in Diabetes)	0.2714	0.0119	<.0001
Age*Diabetes	45-59 years old	Interaction (Age in Diabetes)	0.1303	0.0085	<.0001
Age*Diabetes	75 or older	Interaction (Age in Diabetes)	0.0237	0.0090	0.0084

Parameter	Level	Type	Estimate	Standard Error	p value
Age		Categorical (60-74 is ref)			
15-24 years old		Categorical (60-74 is ref)	0.0045	0.0187	0.8099
25-44 years old		Categorical (60-74 is ref)	-0.2295	0.0077	<.0001
45-59 years old		Categorical (60-74 is ref)	-0.1547	0.0063	<.0001
75 or older		Categorical (60-74 is ref)	-0.0122	0.0063	0.0539
Diabetes		Categorical (0 versus 1)	-0.0633	0.0143	<.0001
Cause of ESRD Missing		Categorical (0 versus 1)	-0.0164	0.0212	0.439
Patient in Nursing Home		Categorical (0 versus 1)	0.5788	0.0053	<.0001
Log of BMI		Continuous	-0.1859	0.0069	<.0001
BMI Missing		Categorical (0 versus 1)	0.1073	0.0097	<.0001
Comorbidity Index		Continuous	0.3624	0.0075	<.0001
Comorbidity Index of 0		Categorical (0 versus 1)	-0.1280	0.0057	<.0001
Comorbidity Index Missing		Categorical (0 versus 1)	-0.1052	0.0274	0.0001
Year	2009	Categorical	-0.0196	0.0048	<.0001
Year	2010	Categorical	-0.0512	0.0048	<.0001
Year	2011	Categorical	0.0169	0.0048	0.0004
Duration_of_ESRD*Diabetes	6 months-1 year	Interaction (Duration of ESRD in Diabetes)	0.0531	0.0173	0.0021
Duration_of_ESRD*Diabetes	1-2 years	Interaction (Duration of ESRD in Diabetes)	0.0794	0.0159	<.0001
Duration_of_ESRD*Diabetes	2-3 years	Interaction (Duration of ESRD in Diabetes)	0.0966	0.0163	<.0001

Duration_of_ESRD*Diabetes	3-5 years	Interaction (Duration of ESRD in Diabetes)	0.0526	0.0155	0.0007
Duration_of_ESRD*Diabetes	5+ years	Interaction (Duration of ESRD in Diabetes)	0.0322	0.0150	0.0322
Age*Diabetes	15-24 years old	Interaction (Age in Diabetes)	0.3414	0.0824	<.0001
Age*Diabetes	25-44 years old	Interaction (Age in Diabetes)	0.2714	0.0119	<.0001
Age*Diabetes	45-59 years old	Interaction (Age in Diabetes)	0.1303	0.0085	<.0001
Age*Diabetes	75 or older	Interaction (Age in Diabetes)	0.0237	0.0090	0.0084

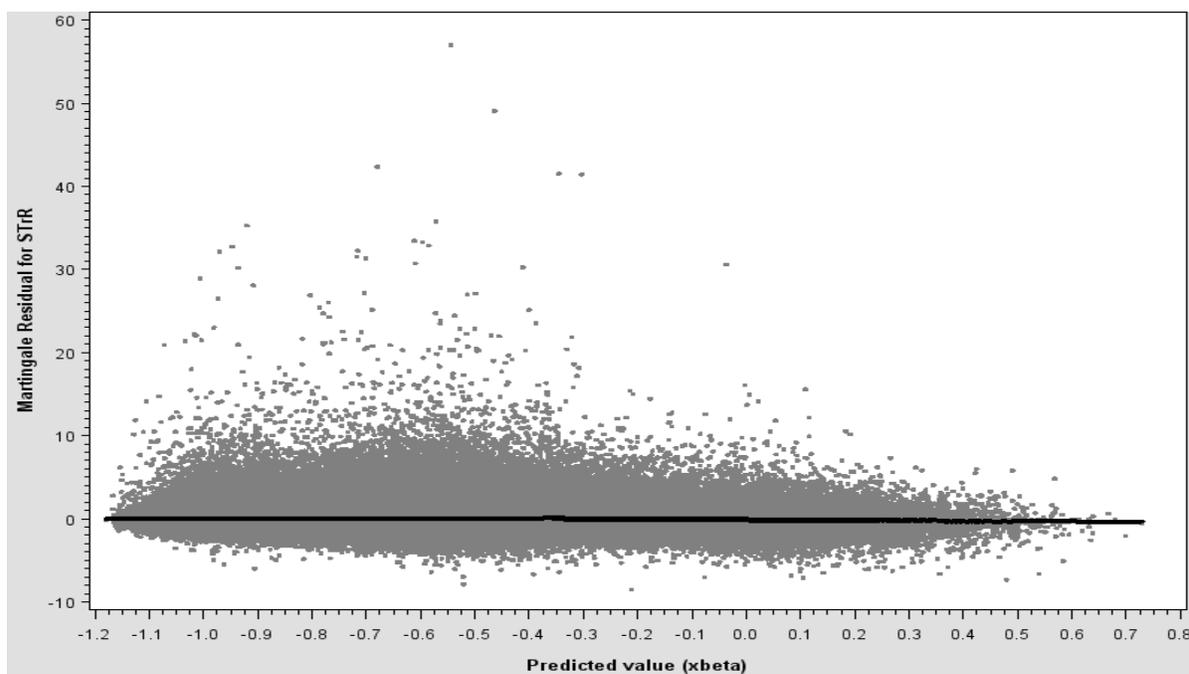
Decile plots (Figure1) shows piecewise linear estimates of the cumulative rates by years since start of ESRD. The plot demonstrates that the risk factors in the model are discriminating well between patients. There is good separation among all 10 groups and the ordering is as predicted by the model (patients predicted to be at lower risk have lower transfusion rates). The absolute differences between the groups is also large with patients predicted to have the highest transfusion rates (line 10) having almost 3 times higher transfusion rates than those predicted to have the lowest rates (line 1).

**Figure1: Decile Plot for Count of Transfusions**



Martingale residual plots were also examined and did not indicate problems with the model fit. The LOESS curve of martingale residuals by predicted value (Figure 2) shows that the mean of the residuals is flat indicating no lack of fit.

**Figure2: Martingale Residuals by Predicted Value with LOESS Curve**



o **Rationale for no adjustment**

2b4.4. *If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment. The three rows above may be deleted if this field is used. Delete row if measure is risk adjusted or if this is a process measure.*

N/A

◆ **Identification of Meaningful Differences in Performance**

o **Data/ sample**

2b5.1 *Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*

Assessment of the STrR was made using data on transfusions among ESRD patients over a period of 2008 to 2011. Refer to section 2a2.1 for the detailed data description.

o **Analytic method**

2b5.2. *Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance*

The STrR is a ratio of the observed number of red blood cell transfusions to the expected number among patients in a facility over a 1-year or 4-year period. The expectation is obtained based on the overall national average rate of transfusions, adjusted for the particular patient mix at the facility under consideration.

In order to classify facilities as having transfusion rates that are better, no different or worse than the national average, we require a method of obtaining a p-value for classification purposes. A p-value assesses the probability that the facility would experience a number of transfusions more extreme than that observed if the null hypothesis

were true; accounting for each facility’s patient mix. To do this, a z-score is first calculated using the estimate and standard error for each facility using the method of generalized estimating equations (GEE; Liang & Zeger, 1986). Specifically, the transfusion rate (or, equivalently: the mean transfusion count, given the exposure) was assumed to follow a multiplicative model and a robust (sandwich) standard error was used. The use of robust standard errors has been advocated for modeling recurrent events (i.e., multiple events per subject) by several previous authors; e.g., Lawless & Nadeau (1995); Lin, Wei, Yang & Ying (2000); Cai & Schaubel (2004). For each facility, the Z-score was computed as the facility’s log(STrR), divided by its standard error. Since log(STrR) is undefined for facilities with 0 transfusions, the Z-score in such cases was computed as (STrR-1), divided by a standard error estimate (sandwich estimator) for STTrR.

To account for the over dispersion in the z-scores, as used in Standardized Hospitalization Ratio (NQF #1463 <http://www.qualityforum.org/QPS/1463>), we use robust estimates of location and scale based on the center of the z-scores (by fitting robust regression on z- scores) and derive normal curves that more closely describes the z-score distribution. This new distribution is referred to as the “empirical null hypothesis” (Efron, 2004) and provide references for assessing the extent to which a given facility’s outcomes are extreme in comparison with other facilities. We then use the mean and standard deviation from the empirical null distribution of the STTrR z-scores to calculate the p-value for classifying facility performance.

References:

- Lin, D.Y., Wei, L.J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society Series B*, 62, 711–730.
- Cai, J. and Schaubel, D.E.. (2004). Marginal means and rates models for multiple-type recurrent event data. *Lifetime Data Analysis*, 10, 121-138.
- Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- Lawless, J.F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37, 158-168.
- Efron, B. (2004). Large scale simultaneous hypothesis testing: the choice of null hypothesis. *J. Amer. Statist. Assoc.*, 99, 96-104.

o **Testing results**

*2b5.3. Results-Provide measure performance results/scores (for example, distribution by quartile, mean, median, SD, etc.); identification of statistically significant and meaningfully differences in performance*

The following Tables 1 and 2 shows how the facilities are flagged for year 2011 and for 2008-2011 (4 year period) respectively, based on the method described above.

**Table1: Classification of Empirical p-value for year 2011**

Year 2011	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Better than expected	4	0.07	4	0.07
As expected	5184	95.88	5188	95.95
Worse than Expected	219	4.05	5407	100

Table 2: Classification of Empirical p-value for 2008-2011

4 year (2008-2011)	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Better than expected	40	0.7	40	0.7
As expected	5374	94.35	5414	95.05
Worse than Expected	282	4.95	5696	100

◆ **Comparability of Multiple Data Sources/Methods**

○ **Data/ sample**

2b6.1. Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included

N/A

○ **Analytic method**

2b6.2. Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure

N/A

○ **Testing results**

2b6.3. Provide statistical results (for example, correlation statistics, comparison of rankings) and assessment of adequacy in the context of norms for the test conducted

N/A

◆ **Disparities in Care**

○ **Stratification**

2c.1. If measure is stratified for disparities, provide stratified results (scores by stratified categories/cohorts)

N/A

- **Rationale for no stratification**

*2c.2. If disparities have been reported/identified, but measure is not specified to detect disparities, please explain.*

Investigations of the STrR by race, sex and ethnicity indicate relatively little variation and no substantial disparities among these groups. Although females are somewhat more likely to receive transfusions than males, analyses showed that a model with race and sex included and a model without these variables yielded very similar results for the facility STrR measure as well as for the parameter estimates for other variables in the model. Hence, stratification was not necessary.

- **Supplemental information**

*2.1. Supplemental testing methodology information: If additional information is available, please indicate where this information can be found: appendix, attachment, or URL*

N/A

## Usability

- ◆ **Public Reporting**

- **Meaningful, understandable and useful**

*3a.1. Use in public reporting---disclosure of performance results to the public at large (If used in a public reporting program, provide name of program(s), locations, Web page URL(s). If not publicly reported in a national or community program, state the reason and plans to achieve public reporting, potential reporting programs or commitments, and timeline, for example, within 3 years of endorsement)*

The STrR may be included on <http://www.medicare.gov/> Dialysis Facility Compare website in the future.

*3a.2. Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. If usefulness was demonstrated (for example, focus, group, cognitive testing) describe the data, method and results.*

CMS has scheduled the STrR to undergo public comment in early 2013, after which CMS will submit the measure for NQF approval. Once the measure has undergone the NQF review process, we plan to include the STrR in the publicly available Dialysis Facility Reports.

- ◆ **Quality Improvement**

- **Meaningful, understandable and useful**

*3b.1. Use in QI (If used in quality improvement program, provide name of program(s), locations, Web page URL(s))*

*3b.2. Provide a rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated (for example, QI, initiative) describe the data, method and results*

This measure is modeled specifically from the KDIGO guidelines and FDA guidance for use of ESAs. Therefore the measure results can act as a useful monitoring tool for facilities' successful adherence to the guidelines. Facilities

that observe increases of the measure over time may be able to identify improvement needs in their anemia management practices.

- o **Other accountability uses**

3.2. *Use for other accountability functions (payment, certification, accreditation) (If used in a public accountability program, provide name of program(s), locations, Web page URL(s)). This row may be deleted if not applicable.*

The STRR may be included in the CMS ESRD Quality Incentive Program in the future.

## Feasibility

- ◆ **How the data elements needed to compute measure score are generated**

4a.1. *How are the data elements needed to compute measure scores generated? State all that apply. Data used in the measure are:*

- o *Generated by and used by health care personnel during the provision of care (for example, blood pressure, lab value, medical condition)*  
*Coded by someone other than person obtaining original information (for example, DRG, ICD-9 codes on claims)*
- o *Abstracted from a record by someone other than person obtaining original information (for example, chart abstraction for quality measure or registry) Other*

Data used in the measure are obtained from Medicare claims generated by and used by health care personnel during the provision of care, i.e. lab values, medical conditions and claims data.

- ◆ **Electronic availability**

4b.1. *Are the data elements needed for the measure as specified available electronically (elements that are needed to compute measure scores are in defined, computer-readable fields)?*

- o *ALL data elements in electronic health records (EHRs)*
- o *ALL data elements in electronic claims*
- o *ALL data elements are in a combination of electronic sources (describe)*
- o *Some data elements are in electronic sources (describe)*
  - *No data elements are in electronic sources*

The data elements needed for the measure as specified are all available electronically.

- ◆ **Susceptibility to inaccuracies, errors, or unintended consequences**

4c.1. *Identify susceptibility to inaccuracies, errors, or unintended consequences of measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results.*

Some lag in transfusion count obtained from hospital claims is expected. If the data are measured too early, the counts may be artificially low. Aside from transfusion data lag, there are no barriers to retrieving the data necessary for the measure, and there are no data availability issues. Burden is minimal for current data because it exists.

◆ **Data collection strategy**

*4d.1. Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues (for example fees for use of proprietary measures)*

The data are from Medicare Part A and B institutional claims.

**Related Measures**

◆ **Harmonization**

*5a.1. If this measure has EITHER the same measure focus OR the same target population as NQF-endorsed measure(s): Are the measure specifications completely harmonized? Is so, describe.*

N/A

◆ **Similar measures**

*5b.1. If this measure has both the same measure focus and the same target population as NQF-endorsed measure(s) or other measures in current use, describe why this measure is superior to existing measures (for example, a more valid or efficient way to measure quality); OR, provide a rationale for the additive value of developing and endorsing an additional measure. (Provide analyses when possible.)*

N/A