NATIONAL QUALITY FORUM

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 2979
Corresponding Measures:
De.2. **Measure Title:** Standardized Transfusion Ratio for Dialysis Facilities
Co.1.1. **Measure Steward:** Centers for Medicare & Medicaid Services
De.3. **Brief Description of Measure:** The risk adjusted facility level transfusion ratio "STrR" is specified for all adult dialysis patients. It is a ratio of the number of eligible red blood cell transfusion events observed in patients dialyzing at a facility, to the number of eligible transfusion events that would be expected under a national norm, after accounting for the patient characteristics within each facility. Eligible transfusions are those that do not have any claims pertaining to the comorbidities identified for exclusion, in the one year look back period prior to each observation window.

This measure is calculated as a ratio, but can also be expressed as a rate.
1b.1. **Developer Rationale:** Safety concerns arising from clinical trials of ESA treatment of anemia of chronic kidney disease (CKD) led to 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 (in 2011) 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, just prior to implementation of the Expanded ESRD Prospective Payment System in 2011.

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 risk for blood transfusion has been reported previously. In addition, the dialysis facility's anemia management process also predicts adverse outcomes associated with anemia management, including need for blood transfusion. These issues are discussed in the Evidence Form included as part of this submission.

S.4. **Numerator Statement:** Number of eligible observed red blood cell transfusion events: An event is defined as the transfer of one or more units of blood or blood products into a recipient's blood stream (code set is provided in the numerator details) among patients dialyzing at the facility during the inclusion episodes of the reporting period. Inclusion episodes are those that do not have any claims pertaining to the comorbidities identified for exclusion, in the one year look back period prior to each observation window.
S.6. **Denominator Statement:** Number of eligible red blood cell transfusion events (as defined in the numerator statement) that would be expected among patients at a facility during the reporting period, given the patient mix at the facility. Inclusion episodes

are those that do not have any claims pertaining to the comorbidities identified for exclusion, in the one year look back period prior to each observation window.
S.8. **Denominator Exclusions:** 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, and sickle cell anemia within one year of their patient time at risk. Since these comorbidities are associated with higher risk of transfusion and require different anemia management practices that the measure is not intended to address, every patient's risk window is modified to have at least 1 year free of claims that contain these exclusion eligible diagnoses.

De.1. **Measure Type:** Outcome
S.17. **Data Source:** Claims, Registry Data
S.20. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Dec 09, 2016 **Most Recent Endorsement Date:** Dec 09, 2016

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** N/A

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
2979_Evidence_form_11022019.docx
1a.1 <u>For Maintenance of Endorsement:</u> Is there new evidence about the measure since the last update/submission?
Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.
No

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
<u>If a COMPOSITE</u> (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.
Safety concerns arising from clinical trials of ESA treatment of anemia of chronic kidney disease (CKD) led to 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 (in 2011) 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, just prior to implementation of the Expanded ESRD Prospective Payment System in 2011.

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 risk for blood transfusion has been reported previously. In addition, the dialysis facility's anemia management process also predicts adverse outcomes associated with anemia management, including need for blood transfusion. These issues are discussed in the Evidence Form included as part of this submission.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*
The STrR is a facility-level measure, comparing the observed number of red blood cell transfusion counts at a facility with the number of transfusions that would be expected under a national norm, after accounting for the patient mix within each facility. Standardized transfusion ratios vary across facilities. The data below show the distribution of STrR using Medicare claims data for 2014-2017.

2014: 6411 facilities, 1.025 mean STrR, 1.010 Standard Error. Facility percentiles: 0.314 (10th), 0.589 (25th), 0.892 (50th), 1.269 (75th), 1.750 (90th).

2015: 6599 facilities, 1.077 mean STrR, 3.782 Standard Error. Facility percentiles: 0.283 (10th), 0.566 (25th), 0.869 (50th), 1.272 (75th), 1.819 (90th).

2016: 6857 facilities, 1.053 mean STrR, 1.855 Standard Error. Facility percentiles: 0.284 (10th), 0.565 (25th), 0.882 (50th), 1.290 (75th), 1.824 (90th)

2017: 7099 facilities, 1.058 mean STrR, 1.211 Standard Error. Facility percentiles: 0.273 (10th), 0.554 (25th), 0.894 (50th), 1.306 (75th), 1.864 (90th)

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*
Analyses of the STrR by race, sex and ethnicity indicate relatively little variation and no disparities substantial to the measure among these groups. Although females are somewhat more likely to receive transfusions than males, analyses showed that a model with variables for 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. The data below show the parameter estimates for the race, sex and ethnicity variables based on a model that included these variables along with other covariates.

Females: 0.152 estimate, 0.004 standard error, <.0001 p-value.
Native American*: -0.044 estimate, 0.020 standard error, 0.031 p-value.
Asian*: -0.174 estimate, 0.011 standard error, <.0001 p-value.

Black*: -0.062 estimate, 0.005 standard error, <.0001 p-value.
Other Race*: -0.043 estimate, 0.034 standard error, 0.202 p-value.
Hispanic#: -0.183 estimate, 0.007 standard error, <.0001 p-value.

*White as reference
Non-Hispanic as reference

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**
N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Renal, Renal : End Stage Renal Disease (ESRD)

De.6. **Non-Condition Specific***(check all the areas that apply):*

De.7. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
N/A

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** 2979_Code_Table_08012019.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
Not an instrument-based measure

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

Two significant revisions were made to the measure in this maintenance submission:

1. For hospital inpatients, the current NQF endorsed STrR relies on a restricted transfusion event identification algorithm. The measure utilizes only those reported transfusion events that include ICD procedure codes, ICD procedure codes with revenue center codes, or value codes*. For the proposed revision to STrR for maintenance, inpatient transfusion events are identified using a broader definition that includes revenue center codes only, ICD procedure codes (alone or with revenue codes), or value codes alone or in combination. The proposed revision will result in identification of a greater number of inpatient transfusion events compared to the currently implemented STrR. In addition, the proposed revision will effectively mitigate a provider coding bias that was exacerbated by the conversion from ICD9 to ICD10 code sets in late CY2015. Identification of outpatient transfusion events is identical in the two STrR versions, as the ICD9 to ICD10 transition does not impact outpatient transfusion claims submission (outpatient claims rely on HCPCS procedure codes instead).

2. The current NQF-endorsed STrR includes all Medicare patients, including those with Medicare Advantage coverage, that meet inclusion criteria based on the presence of Medicare claims "activity" reflected in $900 or greater of outpatient dialysis paid claims in a month or an inpatient claim for a recent inpatient hospitalization. The proposed STrR revision uses similar claims "activity" criteria, but excludes all Medicare Advantage patients' time at risk from both the measure numerator and denominator. This proposed change is being made to mitigate potential bias associated with inclusion of Medicare Advantage patients. The bias derives from the absence of complete outpatient claims data for Medicare Advantage patients, severely limiting the identification of outpatient transfusion events for these individuals, and, eliminating a key source for claims-based exclusion comorbidities.

*ICD9 codes were used prior to October 1, 2015. ICD10 codes have been used since the conversion. We will use the term "ICD codes" throughout the remainder of the documentation.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Number of eligible observed red blood cell transfusion events: An event is defined as the transfer of one or more units of blood or blood products into a recipient's blood stream (code set is provided in the numerator details) among patients dialyzing at the facility during the inclusion episodes of the reporting period. Inclusion episodes are those that do not have any claims pertaining to the comorbidities identified for exclusion, in the one year look back period prior to each observation window.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Transfusion events in the inpatient setting are counted in the following way. The event is identified by presence in a Medicare inpatient claim of the appropriate ICD procedure codes (99.03, 99.04, 30230H1, 30233H1, 30240H1, 30243H1, 30250H1, 30253H1, 30260H1, 30263H1, 30230N1, 30230P1, 30233N1, 30233P1, 30240N1, 30240P1, 30243N1, 30243P1, 30250N1, 30250P1, 30253N1, 30253P1, 30260N1, 30260P1, 30263N1, 30263P1), or revenue center codes (0380, 0381, 0382, 0389, 0390, 0391, 0392, 0399) or value code (37). We only count a single transfusion event for an inpatient claim regardless of the number of transfusion revenue center, procedure and value codes reported so that the number of discrete events counted is the same whether the claim indicates 1 unit of blood or multiple units of blood. This results in a more conservative estimate of blood transfusions from inpatient claims.

Transfusion events are less common in the outpatient setting. Transfusion events in the outpatient setting are counted in the following way. Events derived from outpatient claims are identified by claims with HCPCS code (P9010, P9011, P9016, P9021, P9022, P9038, P9039, P9040, P9051, P9054, P9056, P9058, 36430) with revenue center codes in (0380, 0381, 0382, 0389, 0390, 0391, 0392, 0399) or value code (37). One or more transfusion-related HCPCS codes with at least one transfusion-related revenue center codes, or one or more transfusion-related value codes listed on an outpatient claim are counted as a single transfusion event regardless of

the number of units of blood recorded. In other words, 3 units of blood would be counted as a single transfusion event.

If there are more than one transfusion events identified from inpatient or outpatient claims in the same day, we only count one transfusion event per day.

The detailed procedures to determine unique transfusion events at the claim level are presented in a flow chart in the Appendix (S.19. Calculation Algorithm/Measure Logic Diagram).

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Number of eligible red blood cell transfusion events (as defined in the numerator statement) that would be expected among patients at a facility during the reporting period, given the patient mix at the facility. Inclusion episodes are those that do not have any claims pertaining to the comorbidities identified for exclusion, in the one year look back period prior to each observation window.

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Starting with day 91 after onset of ESRD, a patient is attributed to a facility once the patient has been treated there for the past 60 days and for the following 60 days after transfer to another dialysis facility.

Based on a risk adjustment model for overall national transfusion rates, we compute the expected number of red blood cell transfusion events for each patient attributed to a given facility. The sum of all such expectations over patients in a facility yields the overall expected number of transfusions for the facility given its specific patient mix. This forms the denominator of the measure. This measure is based on Medicare administrative claims and databases and is applied to patients covered by Medicare.

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
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, and sickle cell anemia within one year of their patient time at risk. Since these comorbidities are associated with higher risk of transfusion and require different anemia management practices that the measure is not intended to address, every patient's risk window is modified to have at least 1 year free of claims that contain these exclusion eligible diagnoses.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
We performed multivariate logistic regression demonstrating that a 1-year look back period for the exclusion comorbidities was more predictive of transfusion events compared to longer look back periods. The figure in the appendix describes the inclusion and exclusion period of a hypothetical patient. In the figure included in the exclusion section of the testing form (Sec. 2b2.1), 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 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 a 1-year claim-free period (Inclusion1 and Inclusion2 in the figure). This 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 total transfusion event count because the presence of the exclusion comorbidity claims within the 1-year look back might have increased the risk of transfusion unrelated to dialysis facility anemia management practices. However, T2, which occurs in late 2011 and in Inclusion2 period, will be counted since there is greater than a 1-year gap between this transfusion event and the last claim observed with the exclusion diagnosis.

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

N/A

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
Statistical risk model
If other:

S.12. **Type of score:**
Ratio
If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
The numerator is the observed number of transfusion events for a facility and the denominator for the same facility is the expected number of transfusion events adjusted for patient mix. The measure for a given facility is calculated by dividing the numerator by the denominator. See flowchart for further detail (available in attached appendix).

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.
N/A

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
Specify calculation of response rates to be reported with performance measure results.
N/A

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims, Registry Data

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*
IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.
Data are derived from an extensive national ESRD patient database, which is primarily based on the CMS Consolidated Renal Operations in a Web-enabled Network (CROWN) system. The CROWN data include the Renal Management Information System (REMIS), CROWNWeb facility-reported clinical and administrative data (including CMS-2728 Medical Evidence Form, CMS-2746 Death Notification Form, and CMS-2744 Annual Facility Survey Form data), the historical Standard Information Management System (SIMS) database (formerly maintained by the 18 ESRD Networks until replaced by CROWNWeb in May 2012), the National Vascular Access Improvement Initiative's Fistula First Catheter Last project (in CROWNWeb since May 2012), Medicare dialysis and hospital payment records, transplant data from the Organ Procurement and Transplant Network (OPTN), the Nursing Home Minimum Dataset, the Quality Improvement Evaluation System (QIES) Workbench, which includes data from the Certification and Survey Provider Enhanced Report System (CASPER), the Dialysis Facility Compare (DFC) and the Social Security Death Master File. The database is comprehensive for Medicare patients. Non-Medicare patients are included in all sources except for the Medicare payment records. CROWNWeb provides tracking by dialysis provider and treatment modality for non-Medicare patients. Information on hospitalizations is obtained from Part A Medicare Inpatient Claims Standard Analysis Files (SAFs), and past-year comorbidity is obtained from multiple Part A types (inpatient, home health, hospice, skilled nursing facility claims) and Part B outpatient types of Medicare Claims SAFs.

Information on transfusions is obtained from Medicare Inpatient and Outpatient Claims Standard Analysis Files (SAFs).

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: Dialysis Facility

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2. Validity – See attached Measure Testing Submission Form
2979_Testing_Form_08012019.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.
Yes

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.
Yes

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.
Yes - Updated information is included

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to*

compute the performance measure score are in defined, computer-readable fields) Update this field for **maintenance of endorsement**.
ALL data elements are in defined fields in a combination of electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (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.**
IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.
Data collection is accomplished via Medicare Claims and CROWNWeb, a web-based and electronic batch submission platform maintained and operated by CMS contractors. This maintenance review is in part the result of an ad hoc review request because of a shift in Medicare claims data reporting patterns due to the transition to ICD10 procedure codes that raised questions about the validity of the currently endorsed measure. These issues have been addressed in the revised measure specifications and other parts of this submission.

Measures reported on DFC are reviewed on a regular basis by dialysis facility providers. Review of comments and questions received in the past 3 years for STrR showed only rare instances of concern expressed about inaccurate or missing data.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
N/A

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting

	Dialysis Facility Compare https://www.medicare.gov/dialysisfacilitycompare/ Dialysis Facility Compare https://www.medicare.gov/dialysisfacilitycompare/ Payment Program ESRD Quality Incentive Program https://www.cms.gov/Medicare/Quality-Initiatives-patient-Assessment-Instruments/ESRDQIP/ ESRD Quality Incentive Program https://www.cms.gov/Medicare/Quality-Initiatives-patient-Assessment-Instruments/ESRDQIP/

4a1.1 **For each CURRENT use, checked above (update for <u>maintenance of endorsement</u>), provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

DFC:
Purpose: Dialysis Facility Compare helps patients find detailed information about Medicare-certified dialysis facilities. They can compare the services and the quality of care that facilities provide.
Geographic area: United States
Number of accountable entities: All Medicare-certified dialysis facilities that are eligible for the measure, and have at least 10 patient years at risk (due to public reporting requirements). For the most recent update to Dialysis Facility Compare (October 2019), 6348 facilities had a score reported.
Patients included: All patients who meet the requirements to be included in the measure from included facilities.

QIP:
Purpose: The ESRD QIP will reduce payments to ESRD facilities that do not meet or exceed certain performance standards. The measure was added to the program for PY2018.
Geographic area: United States
Number of accountable entities: All Medicare-certified dialysis facilities that are eligible for the measure, and have at least 10 patient years at risk (due to public reporting requirements). For the most recent QIP report that is publically available (PY 2019), this was 6144 facilities.
Patients included: All patients who meet the requirements to be included in the measure from included facilities.

4a1.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a1.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

4a2.1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
Results of this measure are currently reported on Dialysis Facility Compare and in the ESRD Quality Incentive Program. All Medicare-certified dialysis facilities are eligible for reporting in both programs (approximately 7,000 dialysis facilities). Each program has a

helpdesk and supporting documentation available to assist with interpretation of the measure results.

The measure developer (UM-KECC) produces and distributes the DFC data under contract with CMS. Other CMS contractors calculate and distribute the ESRD QIP measure results.

4a2.1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
For DFC, the results are first reported to facilities via a closed preview period, where facilities can review their data prior to each of the quarterly updates of the public facing Dialysis Facility Compare website. These preview reports are posted on dialysisdata.org, where facilities can also find a detailed Guide to the Quarterly Dialysis Facility Compare Reports and other supporting documentation. Facilities can submit comments/questions about their results at any time, and can request patient lists for their facilities during the specified preview periods.

For the ESRD QIP, results are first reported to facilities via closed preview period on an annual basis; facilities can review their data prior to the results becoming public at the end of the calendar year. These preview reports are posted on qualitynet.org, where facilities can also find supporting documentation and can submit comments/questions about their results.

A measures manual that describes the calculations for both of these programs in detail is published on the CMS website: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/06_MeasuringQuality.html

4a2.2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
For DFC, feedback can be provided any time through contacting the dialysisdata.org helpdesk. Preview periods allow for specific times for facilities review and comment on measure calculations, and provide an opportunity to request a patient list.

For the ESRD QIP, feedback can be provided any time through contacting the QIP helpdesk. Preview periods allow for specific times for facilities review and comment on measure calculations. Comments can also be submitted in response to the Notice of Proposed Rulemaking for each QIP payment year.

For this measure, feedback was also received via an ad hoc review request through the National Quality Forum.

4a2.2.2. **Summarize the feedback obtained from those being measured.**
DFC: Comments received during DFC preview periods tend to be technical in nature, asking for clarification on how the STrR is calculated for particular facilities, including questions about patient assignment and application of exclusion criteria.

QIP: Since the STrR was first proposed in the PY 2018 proposed rule, commenters raised issues related to additional risk adjustment for SDS factors or clinical factors, and question whether the outcome of the measure (transfusions) were attributable to the dialysis facility. Both of these issues are addressed in our submission. Commenters also echoed the concerns raised about measure validity in the ad hoc review request submitted to NQF (see below).

4a2.2.3. **Summarize the feedback obtained from other users**
Ad Hoc Review: Kidney Care Partners have submitted comments for the PY2022 NPRM and to NQF in the form of a request for ad-hoc review of the NQF endorsed STrR arguing that their internal analysis demonstrates that implementation of the currently endorsed version of the STrR results in unintended consequences that can adversely impact facilities when used in CMS' payment and public reporting programs. They base their argument on an analysis using data from 2014-2016 that demonstrates a variable reduction in hospital transfusion events identified with ICD-10 procedure codes (which encompasses data collected beginning in October 2015).

4a2.3. **Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
We have revised the specifications for this measure in response to the ad hoc review request described above. For the proposed revision to STrR, inpatient transfusion events are identified using a broader definition that includes revenue center codes only, ICD procedure codes (alone or with revenue codes), or value codes alone or in combination. The proposed revision will result in identification of a greater number of inpatient transfusion events compared to the currently implemented STrR. In addition, the

proposed revision will effectively mitigate a provider coding bias that was exacerbated by the conversion from ICD9 to ICD10 code sets in late CY2015. Identification of outpatient transfusion events is identical in the two STrR versions, as the ICD9 to 10 transition does not impact outpatient transfusion claims submission.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

CMS is currently reporting this measure on Dialysis Facility Compare (as of January 2014). This measure has also been finalized for the PY2018 QIP. Given that the measure has only been publically reported for a short time, progress on improvement could not be evaluated. We anticipate that public reporting of this measure would improve patient outcomes, given that blood transfusion has been linked to survival indirectly in that transfusions elevate risk of greater exposure to human leukocyte antigens, present in transfused blood, that may increase anti-HLA antibodies in kidney transplant candidates, resulting in reduced access to kidney transplantation for transfused patients. Studies have shown superior patient survival with kidney transplantation compared to chronic dialysis. See 1a.3 for more information.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**

2014: Reference
2015: Estimate = -0.01, p = 0.155, odds ratio = .99
2016: Estimate = -0.04, p = <.0001, odds ratio = .96
2017: Estimate = -0.04, p = <.0001, odds ratio = .96

The modeling results above demonstrate small but significant reductions in inpatient transfusion events for the years 2016-2017 compared to 2014-2015. The predictive model used the broader transfusion event definition that is part of these revised specifications and was adjusted for hospital billing phenotype, hospital size and geographic region.

4b2.2. **Please explain any unexpected benefits from implementation of this measure.**

None

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** 2979_Appendix_11012019.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Helen, Dollar-Maples, Helen.Dollar-Maples@cms.hhs.gov, 410-786-7214-
Co.3 **Measure Developer if different from Measure Steward:** University of Michigan Kidney Epidemiology and Cost Center
Co.4 **Point of Contact:** Jennifer, Sardone, jmsto@med.umich.edu, 734-936-5711-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
This measure was recommended by a Technical Expert Panel in 2012. In this advisory role, the primary duty of the TEP is to suggest candidate measures and related specifications, review any existing measures, and determine if there is sufficient evidence to support the proposed candidate measures. The following were the members of the 2012 TEP that provided their input on the development of this measure.

1.Jeffrey Berns, MD, Professor of Medicine and Pediatrics, University of Pennsylvania School of Medicine
2.Sheila Doss-McQuitty, BSN RN CNN CRA, Nursing Director of Research, Satellite Healthcare, Inc

3.Diana Hlebovy, RN BSN CHN CNN, Clinical Support Specialist, Fresenius Medical Care
4.Robert C Kane, MD FACP*, Acting Deputy Director for Safety, Office of Hematology Oncology Products, CDER
5.Kathe LeBeau, Director of Patient Services and Public Policy, Northeastern Kidney Foundation
6.Harvey Luksenburg, MD*, Chief, Blood Diseases Branch, Division of Blood Diseases and Resources NHLBI
7.Ruth McDonald, MD, Medical Director of Solid Organ Transplant and Ambulatory Services, Seattle Children's Hospital
8.Klemens Meyer, MD, Director of Dialysis Services, Tufts Medical Center
9.John Stivelman, MD, Senior Medical Director and CMO Emeritus, Northwest Kidney Centers

*non-voting

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2016
Ad.3 **Month and Year of most recent revision:** 04, 2016
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?** 04, 2016

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2979
Measure Title: Standardized Transfusion Ratio for Dialysis Facilities
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 11/1/2019

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one

step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome

☒ Outcome: Red Blood Cell Transfusions

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The indication for blood transfusion is usually severe anemia or moderate anemia with recent, active, or anticipated blood loss. Therefore, risk for blood transfusion is dependent on the current degree of anemia (typically measured by hemoglobin concentration or hematocrit%). Management of underlying anemia in chronic dialysis patients is the responsibility of dialysis providers.

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

The Medicare ESRD Program requires Medicare certified dialysis facilities to manage the anemia of CKD as one of their responsibilities under the Conditions for Coverage (1). In addition, the Medicare ESRD Program has included payment for ESAs in dialysis facility reimbursement since 1989. It is notable that inclusion of ESAs in dialysis program payment was associated with a dramatic reduction in the use of blood transfusions in the US chronic dialysis population (2-3).

Recently, reliance on achieved hemoglobin concentration as an indicator of successful anemia management in this population has been de-emphasized and use of other clinically meaningful outcomes, such as transfusion avoidance, have been recommended as alternate measures of anemia management (4-7).

Best dialysis provider practice should include effective anemia management algorithms that focus on 1) prevention and treatment of iron deficiency, inflammation and other causes of ESA resistance, 2) use of the lowest dose of ESAs that achieves an appropriate target hemoglobin that is consistent with FDA guidelines and current best practices, and 3) education of patients, their families and medical providers to avoid unnecessary blood transfusion so that risk of allosensitization is minimized, eliminating or reducing one preventable barrier to successful kidney transplantation.

The decision to transfuse blood is intended to improve or correct the pathophysiologic consequences of severe anemia, defined by achieved hemoglobin or hematocrit%, in a specific clinical context for each patient situation (8). Consensus guidelines in the U.S. and other consensus guidelines defining appropriate use of blood transfusions are based, in large part, on the severity of anemia (9-11). Given the role of hemoglobin as a clinical outcome that defines anemia as well as forms a basis for consensus recommendations regarding use of blood transfusion, it is not surprising that the presence of decreased hemoglobin concentration is a strong predictor of subsequent risk for blood transfusion in multiple settings, including chronic dialysis (12-21). For example, Gilbertson, et al found a nearly four-fold higher risk-adjusted transfusion rate in dialysis patients with achieved hemoglobin <10 gm/dl compared to those with >10 gm/dl hemoglobin. (19) In addition to achieved hemoglobin, other factors related to dialysis facility practices, including the facility's response to their patients achieved hemoglobin, may influence blood transfusion risk in the chronic dialysis population (22, 25). In an observational study recently published by Molony, et al (2016) comparing different facility level titration practices, among patients with hemoglobin <10 and those with hemoglobin>11, they found increased transfusion risk in patients with larger ESA dose reductions and smaller dose escalations, and reduced transfusion risk in patients with larger ESA dose increases and smaller dose reductions (25). The authors reported no clinically meaningful differences in all-cause or cause-specific hospitalization events across groups.

The Food and Drug Administration position defining the primary indication of ESA use in the CKD population is for transfusion avoidance, reflecting the assessment of the relative risks and benefits of ESA use versus blood transfusion. Several historical studies, and one recent research study reviewed by Obrador and Macdougall, document the specific risks of allosensitization after blood transfusion and the potential for transfusion-associated allosensitization to interfere with timely kidney transplantation. (23) A recent analysis demonstrated increased odds ratios for allosensitization associated with transfusion, particularly for men and parous women. That study also demonstrated a 28% reduction in likelihood of transplantation in transfused individuals, based on a multivariate risk-adjusted statistical model. (24)

1. ESRD Facility Conditions for Coverage. https://www.cms.gov/Center/Special-Topic/End-Stage-Renal-Disease-ESRD-Center.html

2. Eschbach et al. Recombinant Human Erythropoietin in Anemic Patients with End-Stage Renal Disease. Results of a Phase III Multicenter Clinical Trial. Annals of Internal Medicine. 1989;111:992-1000.

Study Objective: To determine the effectiveness and safety of recombinant human erythropoietin (rHuEpo).

Patients: Hemodialysis patients (333) with uncomplicated anemia (hematocrit < 0.30). All received rHuEpo intravenously, three times per week at 300 or 150 U/kg body weight, which was then reduced to 75 U/kg and adjusted to maintain the hematocrit at 0.35 ± 0.03 (SD).

Results: The baseline hematocrit (0.223 ± 0.002) increased to 0.35, more than 0.06 over baseline within 12 weeks in 97.4% of patients. Erythrocyte transfusions (1030 within the 6 months before rHuEpo therapy) were eliminated in all patients within 2 months of therapy. Sixty-eight patients with iron overload had a 39% reduction in serum ferritin levels after 6 months of therapy. The median maintenance dose of rHuEpo was 75 U/kg, three times per week (range, 12.5 to 525 U/kg). Nonresponders had complicating causes for anemia: myelofibrosis, osteitis fibrosa, osteomyelitis, and acute or chronic blood loss. Adverse effects included myalgias, 5%; iron deficiency, 43%; increased blood pressure, 35%; and seizures, 5.4%. The creatinine, potassium, and phosphate levels increased slightly but significantly. The platelet count increased slightly but there was no increase in clotting of vascular accesses.

Conclusions: The anemia of hemodialysis patients is corrected by rHuEpo resulting in the elimination of transfusions, reduction in iron overload, and improved quality of life. Iron stores and blood pressure must be monitored and treated to maintain the effectiveness of rHuEpo and to minimize the threat of hypertensive encephalopathy.

3. Powe et al. Early dosing practices and effectiveness of recombinant human erythropoietin. Kidney International, Vol. 43 (1993), pp. 1125—1133.

   Early dosing practices and effectiveness of recombinant human erythropoietin. In a national longitudinal-cohort study of 59,462 end-stage renal disease (ESRD) patients, we examined dosing and effectiveness of erythropoietin (EPO) during the first year of its use in clinical practice(July 1989 through June 1990). In unadjusted and multivariate analyses of Medicare claims data, the mean dose of EPO prescribed was: relatively small and similar for initial and maintenance therapy, 2752 (95% confidence interval 2740 to 2764) and 2668 (95% confidence interval 2654 to 2682) units, respectively; lower when initial therapy was started later (591 units lower in September 1989 and 760 units lower in November 1989 vs. July 1989, P < 0.0001); tower by 135 units during initial therapy and by 116 units during maintenance therapy for females (who weigh less) compared to males (P < 0.001); and lower by 400 units for patients treated in for-profit versus not-for-profit centers. In multivariate analysis: hematocrit response was less and mean maintenance dose was 298 units and 621 units greater for patients whose ESRD was due to multiple myeloma and sickle cell disease, respectively, compared to those with hypertension-related ESRD (P < 0.01); and hematocrit response was logarithmically related to dose [hematocrit =0.97 In (dose), P < 0.0001]. Forty-four percent of patients had a hematocrit $\geq$ 30 after four months of therapy. The percent of patients transfused during three month periods before and after therapy decreased from 20% to 5%, respectively (P < 0.0001). Our results suggest that dosing practices were substantially modified to prescription of smaller and more fixed doses over time, due to the interplay

of clinical concerns and economic forces. They also suggest that the effectiveness of EPO in increasing hematocrit levels and reducing transfusion use in routine clinical practice was less than anticipated based on the experience in clinical trials in part as a result of dosing practices.

4. 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

5. 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. http://www.kdigo.org/clinical_practice_guidelines/pdf/KDIGO-Anemia%20GL.pdf

6. Kliger et al. KDOQI US Commentary on the 2012 KDIGO Clinical Practice Guideline for Anemia in CKD. Am J Kidney Dis. 62(5):849-859.

   The 2012 KDIGO (Kidney Disease: Improving Global Outcomes) Clinical Practice Guideline for Anemia in Chronic Kidney Disease provides clinicians with comprehensive evidence-based recommendations to improve patient care. In this commentary, we review these recommendations and the underlying evidence. Most recommendations are well reasoned. For some, the evidence is unclear and recommendations require some qualification. While the KDIGO guideline stresses the potential risks of intravenous iron therapy, withholding iron might have its own risks. The recommendation to avoid hemoglobin levels falling below 9 g/dL sets a lower bound of “acceptability” that may increase blood transfusion. Given the lack of research supporting the optimal transfusion strategy for end-stage renal disease patients, it is difficult to weigh the risks and benefits of red blood cell transfusion. We find a paucity of evidence that hemoglobin concentration targeted between 11 and 11.5 g/dL is associated with a safety risk. Although the evidence that erythropoiesis-stimulating agent use improves patient quality of life is poor, it is possible that the instruments used to measure quality of life may not be well attuned to the needs of chronic kidney disease or dialysis patients. Our last section focuses specifically on the recommendations to treat anemia in children.

7. Berns, Jeffrey S., Moving Away From Hemoglobin-Based Anemia Performance Measures in Dialysis Patients. Am J Kidney Dis. 2014;64(4):486-488.

   Until recently, dialysis facility quality metrics focused on avoiding low hemoglobin (Hb) concentrations, and financial incentives favored use of erythropoiesis-stimulating agents (ESAs). In many dialysis patients, these practices boosted Hb concentrations to levels that are now considered unnecessary and potentially dangerous. Recent clinical trials have demonstrated that there is little to be gained from, and possible risk in, targeting Hb concentrations > 12-13 g/dL rather than ≤10-11 g/dL.1, 2, 3, 4, 5 Whether the risk is a function of higher Hb concentrations, higher ESA doses, both, or neither remains a matter of debate.6

   International clinical practice guideline recommendations7 and, in the United States,

product labeling by the Food and Drug Administration (FDA) highlight the need to reduce target Hb concentrations and ESA doses. The primary purpose of ESA therapy now is transfusion avoidance. Including the cost of ESAs in the "bundle" as part of the new Prospective Payment System also created a financial disincentive for ESA use. Thus, the conversation about ESA use and Hb concentrations in maintenance hemodialysis patients has shifted from avoiding concentrations that are "too low" to avoiding those that are "too high." However, as predicted, recent data indicate a decline in ESA use and Hb concentrations and an increase in transfusion rates among maintenance hemodialysis patients.8, 9

Recognizing that anemia management performance measures in dialysis units that focused solely on achieved Hb concentration did not improve patient outcomes has prompted interest in moving away from quality improvement metrics that are based on laboratory test results. Instead, interest has shifted toward metrics that reflect outcomes important to patients. In this issue of AJKD, Liu et al10 report a proof-of-concept attempt at developing a dialysis facility–specific standardized transfusion ratio (STfR), a more meaningful anemia quality measure than "What was the Hb concentration last month?" (Developing such a risk-adjusted transfusion metric was a principal recommendation of a Technical Expert Panel meeting hosted by the Arbor Research Collaborative for Health in 2012.11)

8. Whitman, Shreay, Gitlin, van Oijen, & Spiegel. Clinical Factors and the Decision to Transfuse Chronic Dialysis Patients. Clin J Am Soc Nephrol 8: ccc–ccc, 2013. doi: 10.2215/CJN.00160113

Background and objectives: Red blood cell transfusion was previously the principle therapy for anemia in CKD but became less prevalent after the introduction of erythropoiesis-stimulating agents. This study used adaptive choice-based conjoint analysis to identify preferences and predictors of transfusion decision-making in CKD. Design, setting, participants, & measurements: A computerized adaptive choice-based conjoint survey was administered between June and August of 2012 to nephrologists, internists, and hospitalists listed in the American Medical Association Masterfile. The survey quantified the relative importance of 10 patient attributes, including hemoglobin levels, age, occult blood in stool, severity of illness, eligibility for transplant, iron indices, erythropoiesis-stimulating agents, cardiovascular disease, and functional status. Triggers of transfusions in common dialysis scenarios were studied, and based on adaptive choice-based conjoint-derived preferences, relative importance by performing multivariable regression to identify predictors of transfusion preferences was assessed.

Results: A total of 350 providers completed the survey (n=305 nephrologists; mean age=46 years; 21%women).Of 10 attributes assessed, absolute hemoglobin level was the most important driver of transfusions, accounting for 29% of decision-making, followed by functional status (16%) and cardiovascular comorbidities (12%); 92% of providers transfused when hemoglobin was 7.5 g/dl, independent of other factors. In multivariable regression, Veterans Administration providers were more likely to transfuse at 8.0 g/dl (odds ratio, 5.9; 95% confidence interval, 1.9 to 18.4). Although transplant eligibility explained only 5% of decision-making, nephrologists were five times more likely to value it as important compared with non-nephrologists (odds ratio, 5.2; 95% confidence interval, 2.4 to11.1).

Conclusions: Adaptive choice-based conjoint analysis was useful in predicting influences on transfusion decisions. Hemoglobin level, functional status, and cardiovascular comorbidities most strongly influenced transfusion decision-making, but preference variations were observed among subgroups.

9. Carson et al. Red Blood Cell Transfusion: A Clinical Practice Guideline From the AABB. Ann Intern Med. 2012;157:49-58.

   Description: Although approximately 85 million units of red blood cells (RBCs) are transfused annually worldwide, transfusion practices vary widely. The AABB (formerly, the American Association of Blood Banks) developed this guideline to provide clinical recommendations about hemoglobin concentration thresholds and other clinical variables that trigger RBC transfusions in hemodynamically stable adults and children. Methods: These guidelines are based on a systematic review of randomized clinical trials evaluating transfusion thresholds. We performed a literature search from 1950 to February 2011 with no language restrictions. We examined the proportion of patients who received any RBC transfusion and the number of RBC units transfused to describe the effect of restrictive transfusion strategies on RBC use. To determine the clinical consequences of restrictive transfusion strategies, we examined overall mortality, nonfatal myocardial infarction, cardiac events, pulmonary edema, stroke, thromboembolism, renal failure, infection, hemorrhage, mental confusion, functional recovery, and length of hospital stay.
   Recommendation 1: The AABB recommends adhering to a restrictive transfusion strategy (7 to 8 g/dL) in hospitalized, stable patients (Grade: strong recommendation; high-quality evidence).
   Recommendation 2: The AABB suggests adhering to a restrictive strategy in hospitalized patients with preexisting cardiovascular disease and considering transfusion for patients with symptoms or a hemoglobin level of 8 g/dL or less (Grade: weak recommendation; moderate-quality evidence).
   Recommendation 3: The AABB cannot recommend for or against a liberal or restrictive transfusion threshold for hospitalized, hemodynamically stable patients with the acute coronary syndrome (Grade: uncertain recommendation; very low-quality evidence).
   Recommendation 4: The AABB suggests that transfusion decisions be influenced by symptoms as well as hemoglobin concentration (Grade: weak recommendation; low-quality evidence).

10. American Society of Anesthesiologists Task Force on Perioperative Blood Transfusion and Adjuvant Therapies. Practice guidelines for perioperative blood transfusion and adjuvant therapies: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Blood Transfusion and Adjuvant Therapies. *Anesthesiology*. 2006;105:198–208.

11. Munoz et al. “Fit to fly”; overcoming barriers to preoperative haemoglobin optimization in surgical patients. Br J Anaesth. 2015 Jul;115(1):15-24.

    In major surgery, the implementation of multidisciplinary, multimodal and individualized strategies, collectively termed Patient Blood Management, aims to identify modifiable risks and optimise patients' own physiology with the ultimate goal of improving outcomes. Among the various strategies utilized in Patient Blood Management, timely detection and management of preoperative anaemia is most important, as it is in itself a

risk factor for worse clinical outcome, but also one of the strongest predisposing factors for perioperative allogeneic blood transfusion, which in turn increases postoperative morbidity, mortality and costs. However, preoperative anaemia is still frequently ignored, with indiscriminate allogeneic blood transfusion used as a 'quick fix'. Consistent with reported evidence from other medical specialties, this imprudent practice continues to be endorsed by non-evidence based misconceptions, which constitute serious barriers for a wider implementation of preoperative haemoglobin optimisation. We have reviewed a number of these misconceptions, which we unanimously consider should be promptly abandoned by health care providers and replaced by evidence-based strategies such as detection, diagnosis and proper treatment of preoperative anaemia. We believe that this approach to preoperative anaemia management may be a viable, cost-effective strategy that is beneficial both for patients, with improved clinical outcomes, and for health systems, with more efficient use of finite health care resources.

12. Dunne, Malone, Tracy, Gannon, and Napolitano. Perioperative Anemia: An Independent Risk Factor for Infection, Mortality, and Resource Utilization in Surgery. Journal of Surgical Research 102, 237-244 (2002)

    Background. Previous studies on patients with hip fractures and in patients with colorectal cancer have documented that perioperative transfusion is associated with a significant increase in postoperative infection rate. Therefore, we sought to investigate the incidence of preoperative and postoperative anemia in noncardiac surgical patients and to determine if transfusion is an independent risk factor for infection and adverse outcome postoperatively.
    Methods. Prospective data from the National Veterans Administration Surgical Quality Improvement Program (NSQIP) was collected on 6301 noncardiac surgical patients at the Veterans Affairs Maryland Healthcare System from 1995 to 2000.
    Results. The mean age of the study cohort was 61 6 13. Descriptive data revealed 95% were male, 44% used tobacco, 19% were diabetic, 9% had COPD, 9% used alcohol, 3% used steroids, 1.7% had a diagnosis of cancer, and 1.2% had ascites. Preoperative anemia (hematocrit less than 36) was found in 33.9% and postoperative anemia was found in 84.1% of the study cohort. In the postoperative period, 32.5% of patients had a hematocrit of 26±30, and 26.5% had a hematocrit of 21±25. Mean units of blood transfused in the perioperative period ranged from 0.1 6 0.9 in patients without anemia to 2.7 6 2.9 in those with anemia. Incidence of pneumonia increased from 2.6 to 5% with increasing degree of anemia. Multiple logistic regression analysis documented that low preoperative hematocrit, low postoperative hematocrit, and increased blood transfusion rates were associated with increased mortality ($P < 0.01$), increased postoperative pneumonia ($P < 0.05$), and increased hospital length of stay ($P < 0.05$).
    Conclusion. There is a high incidence of preoperative and postoperative anemia in surgical patients, with a coincident increase in blood utilization. These factors are associated with increased risk for perioperative infection and adverse outcome (mortality) in surgical patients. Consideration should be given to preoperative diagnosis and correction of anemia with iron, vitamin B12, folate supplementation, or administration of recombinant human erythropoietin.

13. Covin R, O'Brien M, Grunwald G, Brimhall B, Sethi G, Walczak S, Reiquam W, Rajagopalan C, Shroyer AL Factors affecting transfusion of fresh frozen plasma, platelets, and red blood cells during elective coronary artery bypass graft surgery. Arch Pathol Lab Med. 2003 Apr;127(4):415-23.

CONTEXT: The ability to predict the use of blood components during surgery will improve the blood bank's ability to provide efficient service. OBJECTIVE: Develop prediction models using preoperative risk factors to assess blood component usage during elective coronary artery bypass graft surgery (CABG). DESIGN: Eighty-three preoperative, multidimensional risk variables were evaluated for patients undergoing elective CABG-only surgery. MAIN OUTCOMES MEASURES: The study endpoints included transfusion of fresh frozen plasma (FFP), platelets, and red blood cells (RBC). Multivariate logistic regression models were built to assess the predictors related to each of these endpoints. SETTING: Department of Veterans Affairs (VA) health care system. PATIENTS: Records for 3034 patients undergoing elective CABG-only procedures; 1033 patients received a blood component transfusion during CABG. RESULTS: Previous heart surgery and decreased ejection fraction were significant predictors of transfusion for all blood components. Platelet count was predictive of platelet transfusion and FFP utilization. Baseline hemoglobin was a predictive factor for more than 2 units of RBC. Some significant hospital variation was noted beyond that predicted by patient risk factors alone. CONCLUSIONS: Prediction models based on preoperative variables may facilitate blood component management for patients undergoing elective CABG. Algorithms are available to predict transfusion resources to assist blood banks in improving responsiveness to clinical needs. Predictors for use of each blood component may be identified prior to elective CABG for VA patients.

14. Jans et al. Role of preoperative anemia for risk of transfusion and postoperative morbidity in fast-track hip and knee arthroplasty. Transfusion. 2014 Mar;54(3):717-26.

BACKGROUND: Preoperative anemia has been associated with increased risk of allogeneic blood transfusion and postoperative morbidity and mortality. The prevalence of preoperative anemia and its association with postoperative outcomes has not previously been reported in relation to fast-track elective total hip arthroplasty (THA) and total knee arthroplasty (TKA). We aimed to evaluate the prevalence of preoperative anemia in elective fast-track THA and TKA and its association with risk of perioperative transfusion, prolonged length of hospital stay (LOS), and postoperative readmission. STUDY DESIGN AND METHODS: This was a prospective observational database study with data obtained from six high-volume Danish fast-track surgical centers. Preoperative hemoglobin and patient demographics were collected prospectively using questionnaires while outcome and transfusion data were collected using national databases and patient charts. Adjusted risk estimates for transfusion, prolonged LOS, and all-cause readmission according to preoperative anemia status were obtained by multivariate logistic regression. RESULTS: A total of 5.165 THA or TKA procedures were included with a mean patient age of 67 ± 11 years and a median LOS of 2 (interquartile range, 2-3) days. A total of 662 patients (12.8%) had preoperative anemia according to World Health Organization classification. Preoperative anemia was associated with increased risk of receiving transfusion during admission (odds ratio [OR], 4.7; 95% confidence interval [CI], 3.8-5.8), increased risk of readmission within 90 days from

surgery (OR, 1.4; 95% CI, 1.1-1.9), and increased risk of LOS of more than 5 days (OR, 2.5; 95% CI, 1.9-3.4) after adjustment for preoperative patient-related risk factors. CONCLUSION: Preoperative anemia in elective fast-track THA and TKA is independently associated with transfusion and increased postoperative morbidity, supporting the need for preoperative evaluation and treatment.

15. Saleh et al. Allogenic Blood Transfusion Following Total Hip Arthroplasty: Results from the Nationwide Inpatient Sample, 2000 to 2009. J Bone Joint Surg Am. 2014;96:e155(1-10)

    Background: The large-scale utilization of allogenic blood transfusion and its associated outcomes have been described in critically ill patients and those undergoing high-risk cardiac surgery but not in patients undergoing elective total hip arthroplasty. The objective of this study was to determine the trends in utilization and outcomes of allogenic blood transfusion in patients undergoing primary total hip arthroplasty in the United States from 2000 to 2009.
    Methods: An observational cohort of 2,087,423 patients who underwent primary total hip arthroplasty from 2000 to 2009 was identified in the Nationwide Inpatient Sample. International Classification of Diseases, Ninth Revision, Clinical Modification procedure codes 99.03 and 99.04 were used to identify patients who received allogenic blood products during their hospital stay. Risk factors for allogenic transfusions were identified with use of multivariable logistic regression models. We used propensity score matching to estimate the adjusted association between transfusion and surgical outcomes.
    Results: The rate of allogenic blood transfusion increased from 11.8% in 2000 to 19.0% in 2009. Patient-related risk factors for receiving an allogenic blood transfusion include an older age, female sex, black race, and Medicaid insurance. Hospital-related risk factors include rural location, smaller size, and non-academic status. After adjusting for confounders, allogenic blood transfusion was associated with a longer hospital stay (0.58 ± 0.02 day; $p < 0.001$), increased costs ($1731 ± $49 [in 2009 U.S. dollars]; $p < 0.001$), increased rate of discharge to an inpatient facility (odds ratio, 1.28; 95% confidence interval, 1.26 to 1.31), and worse surgical and medical outcomes. In-hospital mortality was not affected by allogenic blood transfusion (odds ratio, 0.97; 95% confidence interval, 0.77 to 1.21).
    Conclusions: The increase in allogenic blood transfusion among total hip arthroplasty patients is concerning considering the associated increase in surgical complications and adverse events. The risk factors for transfusion and its impact on costs and inpatient outcomes can potentially be used to enhance patient care through optimizing preoperative discussions and effective utilization of blood-conservation methods.
    Level of Evidence: Therapeutic Level IV. See Instructions for Authors for a complete description of levels of evidence.

16. Ejaz, Spolverato, Kim, Frank, and Pawlik. Variations in triggers and use of perioperative blood transfusions in major gastrointestinal surgery. Br. J. Surg. 2014 Oct;101(11):1424-33.

    BACKGROUND: The decision to perform intraoperative blood transfusion is subject to a variety of clinical and laboratory factors. This study examined variation in haemoglobin (Hb) triggers and overall utilization of intraoperative blood transfusion, as well the impact of transfusion on perioperative outcomes. METHODS: The study included all

patients who underwent pancreatic, hepatic or colorectal resection between 2010 and 2013 at Johns Hopkins Hospital, Baltimore, Maryland. Data on Hb levels that triggered an intraoperative or postoperative transfusion and overall perioperative blood utilization were obtained and analysed. RESULTS: Intraoperative transfusion was employed in 437 (15·6 per cent) of the 2806 patients identified. Older patients (odds ratio (OR) 1·68), patients with multiple co-morbidities (Charlson co-morbidity score 4 or above; OR 1·66) and those with a lower preoperative Hb level (OR 4·95) were at increased risk of intraoperative blood transfusion (all P < 0·001). The Hb level employed to trigger transfusion varied by sex, race and service (all P < 0·001). A total of 105 patients (24·0 per cent of patients transfused) had an intraoperative transfusion with a liberal Hb trigger (10 g/dl or more); the majority of these patients (78; 74·3 per cent) did not require any additional postoperative transfusion. Patients who received an intraoperative transfusion were at greater risk of perioperative complications (OR 1·55; P = 0·002), although patients transfused with a restrictive Hb trigger (less than 10 g/dl) showed no increased risk of perioperative morbidity compared with those transfused with a liberal Hb trigger (OR 1·22; P = 0·514). CONCLUSION: Use of perioperative blood transfusion varies among surgeons and type of operation. Nearly one in four patients received a blood transfusion with a liberal intraoperative transfusion Hb trigger of 10 g/dl or more. Intraoperative blood transfusion was associated with higher risk of perioperative morbidity.

17. Foley, Curtis, & Parfrey. Hemoglobin Targets and Blood Transfusions in Hemodialysis Patients without Symptomatic Cardiac Disease Receiving Erythropoietin Therapy. Clin J Am Soc Nephrol 3: 1669–1675, 2008. doi: 10.2215/CJN.02100508 .

    Background and objectives: Optimal hemoglobin targets for chronic kidney disease patients receiving erythropoiesis-stimulating agents remain controversial. The effects of different hemoglobin targets on blood transfusion requirements have not been well characterized, despite their relevance to clinical decision-making.
    Design, setting, participants, & measurements: Five hundred ninety-six incident hemodialysis patients without symptomatic cardiac disease were randomly assigned to hemoglobin targets of 9.5 to 11.5 g/dl or 13.5 to 14.5 g/dl for 96 wk using epoetin alfa as primary therapy and changes in left ventricular structure as the primary outcome (previously reported). Patients were masked to treatment assignment. Blood transfusion data were prospectively collected at 4-wk intervals.
    Results: The mean age and prior duration of dialysis therapy of the study population were 50.8 and 0.8 yr, respectively. Previously reported mortality was similar in low and high-target subjects, at 4.7 (95% confidence interval 3.0, 7.3) and 3.1 (1.8, 5.4) per hundred patient years, respectively. Transfusion rates were 0.66 (0.59, 0.74) units of blood per year in low and 0.26 (0.22, 0.32) in high-target subjects (P < 0.0001). Hemoglobin level at transfusion (7.7 [7.5, 7.9]) versus 8.1 [7.6, 8.5] g/dl) were similar with both groups. High hemoglobin target was a significant predictor of time to first transfusion independent of baseline associations (hazard ratio 0.42; 95% confidence interval 0.26 – 0.67). Conclusions: In hemodialysis patients with comparatively low mortality risks, normal hemoglobin targets may reduce the need for transfusions.

18. Hirth, Turenne, Wilk et al. Blood transfusion practices in dialysis patients in a dynamic regulatory

environment. Am J Kidney Dis. 2014 Oct;64(4):616-21. doi: 10.1053/j.ajkd.2014.01.011. Epub 2014 Feb 19.

BACKGROUND: In 2011, Medicare implemented a prospective payment system (PPS) covering an expanded bundle of services that excluded blood transfusions. This led to concern about inappropriate substitution of transfusions for other anemia management methods.
STUDY DESIGN: Medicare claims were used to calculate transfusion rates among dialysis patients pre- and post-PPS. Linear probability regressions adjusted transfusion trends for patient characteristics.
SETTING & PARTICIPANTS: Dialysis patients for whom Medicare was the primary payer between 2008 and 2012.
PREDICTOR: Pre-PPS (2008-2010) versus post-PPS (2011-2012).
OUTCOMES & MEASUREMENTS: Monthly and annual probability of receiving one or more blood transfusions.
RESULTS: Monthly rates of one or more transfusions varied from 3.8%-4.8% and tended to be lowest in 2010. Annual rates of transfusion events per patient were ~10% higher in relative terms post-PPS, but the absolute magnitude of the increase was modest (~0.05 events/patient). A larger proportion received 4 or more transfusions (3.3% in 2011 and 2012 vs 2.7%-2.8% in prior years). Controlling for patient characteristics, the monthly probability of receiving a transfusion was significantly higher post-PPS ($\beta = 0.0034$; $P < 0.001$), representing an ~7% relative increase. Transfusions were more likely for females and patients with more comorbid conditions and less likely for blacks both pre- and post-PPS.
LIMITATIONS: Possible underidentification of transfusions in the Medicare claims, particularly in the inpatient setting. Also, we do not observe which patients might be appropriate candidates for kidney transplantation.
CONCLUSIONS: Transfusion rates increased post-PPS, but these increases were modest in both absolute and relative terms. The largest increase occurred for patients already receiving several transfusions. Although these findings may reduce concerns regarding the impact of Medicare's PPS on inappropriate transfusions that impair access to kidney transplantation or stress blood bank resources, transfusions should continue to be monitored.

19. Gilbertson, Monda, Bradbury & Collins. RBC Transfusions Among Hemodialysis Patients (1999-2010): Influence of Hemoglobin Concentrations Below 10 g/dL. Am J Kidney Dis. 2013; Volume 62 , Issue 5 , 919 - 928

Background: Changes in anemia management over the past decade have produced downward shifts in hemoglobin concentrations. We aimed to examine the effect on use of red blood cell (RBC) transfusions.
Study Design: Retrospective cohort study.
Setting & Participants: We identified point prevalent Medicare hemodialysis patients as of January 1 of each year (1999-2010) and categorized them based on 3-month (April to June) mean hemoglobin levels (10 or 10 g/dL) in each year.
Predictors: Hemoglobin patterns over time and clinical profiles based on achieved hemoglobin concentrations.
Outcomes: RBC transfusion use. Measurements: We used negative binomial modeling to

examine the effect of hemoglobin level 10 g/dL on transfusion use, adjusting for case-mix differences.
Results: Proportions of patients with mean hemoglobin levels10 g/dL decreased from 10% (1999) to4% (2005), but began increasing after 2006 and reached 6% by 2010. Accounting for case-mix differences, transfusion rates remained relatively constant at approximately 7.9 per 100 person-months for patients with hemoglobin levels 10 g/dL and 2 per 100 person-months for patients with hemoglobin levels 10 g/dL.
Patients with average hemoglobin levels 10 g/dL were more likely to receive transfusions (risk ratio, 2.2; 95% CI, 2.1-2.2) even after adjustment; the risk ratio doubled if hemoglobin levels remained 10 g/dL for 6 months (4.4; 95% CI, 3.7-5.2).
Limitations: Limited in generalizability to patients with Medicare as primary payer; residual confounding from factors such as frailty and chronic inflammation cannot be excluded; categorizing patients based on an average of 3 outpatient hemoglobin measurements may introduce some misclassification.
Conclusions: Risk of transfusion increases substantially with hemoglobin concentrations 10 g/dL; risk appears to be independent of other clinical factors. If anemia management patterns shift toward lower hemoglobin concentrations, RBC transfusion use likely will increase in dialysis patients.

20. Collins et al. Effect of Facility-Level Hemoglobin Concentration on Dialysis Patient Risk of Transfusion. Am J Kidney Dis. 2014; 63(6):997-1006.

    Background: Changes in anemia management practices due to concerns about erythropoiesis-stimulating agent safety and Medicare payment changes may increase patient risk of transfusion. We examined anemia management trends in hemodialysis patients and risk of red blood cell (RBC) transfusion according to dialysis facility–level hemoglobin concentration.
    Study Design: Retrospective follow-up study; 6-month study period (January to June), 3-month exposure/follow-up.
    Setting & Participants: For each year in 2007-2011, annual cohorts of point-prevalent Medicare primary payer patients receiving hemodialysis on January 1with one or more hemoglobin measurements during the study period. Annual cohorts averaged 170,000 patients, with 130,000 patients and 3,100 facilities for the risk analysis.
    Predictor: Percentage of facility patient-months with hemoglobin level, 10 g/dL.
    Outcome: Patient-level RBC transfusion rates.
    Measurements: Monthly epoetin alfa and intravenous iron doses, mean hemoglobin levels, and RBC transfusion rates; percentage of facility patient-months with hemoglobin levels, 10 g/dL (exposure) and patient-level RBC transfusion rates (follow-up).
    Results: Percentages of patients with hemoglobin levels, 10 g/dL increased every year from 2007 (6%) to 2011 (w11%). Epoetin alfa doses, iron doses, and transfusion rates remained relatively stable through 2010 and changed in 2011. Median monthly epoetin alfa and iron doses decreased 25% and 43.8%, respectively, and monthly transfusion rates increased from 2.8% to 3.2% in 2011, a 14.3% increase. Patients in facilities with the highest prevalence of hemoglobin levels , 10 g/dL over 3 months were at w30% elevated risk of receiving RBC transfusions within the next 3 months (relative risk, 1.28; 95% CI, 1.22-1.34).
    Limitations: Possibly incomplete claims data; smaller units excluded; hemoglobin levels reported monthly for patients receiving epoetin alfa; transfusions usually not

administered in dialysis units.
Conclusions: Dialysis facility treatment practices, as assessed by percentage of patient-months with hemoglobin levels , 10 g/dL over 3 months, were associated significantly with risk of transfusions in the next 3 months for all patients in the facility, regardless of patient case-mix.

21. Cappell et al. Red blood cell (RBC) transfusion rates among US chronic dialysis patients during changes to Medicare end-stage renal disease (ESRD) reimbursement systems and erythropoiesis stimulating agent (ESA) labels. BMC Nephrology 2014, 15:116.

    Background: Several major ESRD-related regulatory and reimbursement changes were introduced in the United States in 2011. In several large, national datasets, these changes have been associated with decreases in erythropoiesis stimulating agent (ESA) utilization and hemoglobin concentrations in the ESRD population, as well as an increase in the use of red blood cell (RBC) transfusions in this population. Our objective was to examine the use of RBC transfusion before and after the regulatory and reimbursement changes implemented in 2011 in a prevalent population of chronic dialysis patients in a large national claims database.
    Methods: Patients in the Truven Health MarketScan Commercial and Medicare Databases with evidence of chronic dialysis were selected for the study. The proportion of chronic dialysis patients who received any RBC transfusion and RBC transfusion event rates per 100 patient-months were calculated in each month from January 1, 2007 to March 31,2012. The results were analyzed overall and stratified by primary health insurance payer (commercial payer or Medicare).
    Results: Overall, the percent of chronic dialysis patients with RBC transfusion and RBC transfusion event rates per 100 patient-months increased between January 2007 and March 2012. When stratified by primary health insurance payer, it appears that the increase was driven by the primary Medicare insurance population. While the percent of patients with RBC transfusion and RBC transfusion event rates did not increase in the commercially insured population between 2007 and 2012 they did increase in the primary Medicare insurance population; the majority of the increase occurred in 2011 during the same time frame as the ESRD-related regulatory and reimbursement changes.
    Conclusions: The regulatory and reimbursement changes implemented in 2011 may have contributed to an increase in the use of RBC transfusions in chronic dialysis patients in the MarketScan dataset who were covered by Medicare plus Medicare supplemental insurance.

22. House AA, Pham B, Pagé DE. Transfusion and recombinant human erythropoietin requirements differ between dialysis modalities. Nephrol Dial Transplant. 1998 Jul;13(7):1763-9.

    BACKGROUND: Before the routine use of recombinant human erythropoietin (rHuEpo), patients dialysed by peritoneal dialysis (PD) received fewer blood transfusions than patients on haemodialysis (HD). We compared transfusion practices in these groups now that the use of rHuEpo has become standard, while controlling for variables known to influence anaemia of end-stage renal disease (ESRD). Maintenance rHuEpo doses were also compared. METHODS: Data were examined for 157 HD and 126 PD patients

during a 2-year period. Potential confounders included age, gender, albumin, iron deficiency, parathyroid hormone (PTH), underlying renal disease, comorbid illness, renal transplant, dialysis adequacy and duration. An intent-to-treat analysis was used, with sensitivity analyses to account for change in treatment and transplant. RESULTS: Mean haemoglobin (Hb) was not different (10.47 g/dl for HD, 10.71 g/dl for PD; P = 0.45). Mean monthly transfusion rate was higher for HD (0.47 units per month vs 0.19; P < 0.01). More HD patients received at least one transfusion (52.9 vs 40.9%; P < 0.01). The maintenance rHuEpo dose was higher for HD (7370 U/week vs 5790 U/week; P = 0.01). The only factors associated with risk of being transfused were dialysis duration and mode of dialysis (less risk for PD, odds-ratio 0.57; 95% confidence interval 0.35-0.92). CONCLUSIONS: Despite the routine use of rHuEpo, HD patients received more blood and rHuEpo than PD patients to achieve the same Hb. No patient factors were identified to account for this difference. The use of fewer transfusions and less rHuEpo in PD represents an advantage over HD in terms of both cost and safety.

23. Obrador and Macdougall. Effect of Red Cell Transfusions on Future Kidney Transplantation. Clin J Am Soc Nephrol 8: 852–860, 2013.

    Red cell transfusions, erythropoiesis-stimulating agents (ESAs), and intravenous iron therapy all have a place in the treatment of anemia associated with CKD. Their relative merits and uses are subject to many clinical and nonclinical factors. New concerns associated with the use of ESA therapy make it likely that the use of blood transfusions will increase, refueling previous debates about their associated risks. Data on whether red cell transfusions increase sensitization to HLA antigens, rendering subsequent transplantation more problematic, are mainly derived from older literature. Older data suggested that women were more at risk of HLA sensitization than men, particularly those with previous multiple pregnancies, although recent U.S. Renal Data System data have challenged this. HLA sensitization prolongs the waiting time for transplantation and reduces graft survival. Leukocyte depletion of red cells does not appear to reduce the risk of HLA sensitization. This review summarizes much of the data on these issues, as well as highlighting the need for further research on the potential risks for blood transfusion in patients with CKD.

24. Ibrahim, et al. Blood transfusions in kidney transplant candidates are common and associated with adverse outcomes. Clin Transplant 2011: 25: 653–659.

    Surprisingly, there are no data regarding transfusion frequency, factors associated with transfusion administration in patients on the kidney transplant waiting list, or transfusion impact on graft and recipient outcomes. We used United States Renal Data System data to identify 43 025 patients added to the waiting list in 1999–2004 and followed through 2006 to assess the relative risk of post-listing transfusions. In 69 991 patients who underwent transplants during the same time period, we assessed the association between pre-transplant transfusions and level of panel-reactive antibody (PRA) at the time of transplant, and associations between PRA and patient outcomes. The three-yr cumulative incidence of transfusions was 26% for patients added to the waiting list in 1999, rising to 30% in

2004. Post-listing transfusions were associated with a 28% decreased likelihood of undergoing transplant, and a more than fourfold increased risk of death. There was a graded association between percent PRA at the time of transplant and adjusted risk of death-censored graft failure, death with function, and the combined event of graft failure and death. These data demonstrate that transfusions remain common and confirm the adverse association between transfusions and PRA, and high PRA and inferior graft and patient outcomes.

25. Molony, et al. Effects of epoetin alfa titration practices, implemented after changes to product labeling, on hemoglobin levels, transfusion use, and hospitalization rates. Am J Kidney Dis 2016: epub before print (published online March 12, 2016).

Background: Little is known about epoetin alfa (EPO) dosing at dialysis centers after implementation of the US Medicare prospective payment system and revision of the EPO label in 2011.
Study Design: Retrospective cohort study.
Setting & Participants: Approximately 412,000 adult hemodialysis patients with Medicare Parts A and B as primary payer in 2009 to 2012 to describe EPO dosing and hemoglobin patterns; of these, about 70,000 patients clustered in about 1,300 dialysis facilities to evaluate facility-level EPO titration practices and patient level outcomes in 2012.
Predictor: Facility EPO titration practices when hemoglobin levels were ,10 and .11 g/dL (grouped treatment variable) determined from monthly EPO dosing and hemoglobin level patterns.
Outcomes: Patient mean hemoglobin levels, red blood cell transfusion rates, and all-cause and cause specific hospitalization rates using a facility-based analysis.
Measurements: Monthly EPO dose and hemoglobin level, red blood cell transfusion rates, and all-cause and cause-specific hospitalization rates.
Results: Monthly EPO doses declined across all hemoglobin levels, with the greatest decline in patients with hemoglobin levels , 10 g/dL (July-October 2011). In 2012, nine distinct facility titration practices were identified. Across groups, mean hemoglobin levels differed slightly (10.5-10.8 g/dL) but within-patient hemoglobin standard deviations were similar (w0.68 g/dL). Patients at facilities implementing greater dose reductions and smaller dose escalations had lower hemoglobin levels and higher transfusion rates. In contrast, patients at facilities that implemented greater dose escalations (and large or small dose reductions) had higher hemoglobin levels and lower transfusion rates. There were no clinically meaningful differences in all-cause or cause-specific hospitalization events across groups.
Limitations: Possibly incomplete claims data; excluded small facilities and those without consistent titration patterns; hemoglobin levels reported monthly; inferred facility practice from observed dosing.
Conclusions: Following prospective payment system implementation and labeling revisions, EPO doses declined significantly. Under the new label, facility EPO titration practices were associated with mean hemoglobin levels (but not standard deviations) and transfusion use, but not hospitalization rates.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	

Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

_____

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 2979
Measure Title: Standardized Transfusion Ratio for Dialysis Facilities
Date of Submission: 8/1/2019
Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☒ registry	☒ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Data are derived from an extensive national ESRD patient database, which is primarily based on the CMS Consolidated Renal Operations in a Web-enabled Network (CROWN) system. The CROWN data include the Renal Management Information System (REMIS), CROWNWeb facility-reported clinical and administrative data (including CMS-2728 Medical Evidence Form, CMS-2746 Death Notification Form, and CMS-2744 Annual Facility Survey Form data), the historical Standard Information Management System (SIMS) database (formerly maintained by the 18 ESRD Networks until replaced by CROWNWeb in May 2012), the National Vascular Access Improvement Initiative's Fistula First Catheter Last project (in CROWNWeb since May 2012), Medicare dialysis and hospital payment records, transplant data from the

Organ Procurement and Transplant Network (OPTN), the Nursing Home Minimum Dataset, the Quality Improvement Evaluation System (QIES) Workbench, which includes data from the Certification and Survey Provider Enhanced Report System (CASPER), Dialysis Facility Compare (DFC), and the Social Security Death Master File. The database is comprehensive for Medicare patients. Non-Medicare patients are included in all sources except for the Medicare payment records. CROWNWeb provides tracking by dialysis provider and treatment modality for non-Medicare patients. Information on hospitalizations is obtained from Part A Medicare Inpatient Claims Standard Analysis Files (SAFs), and past-year comorbidity is obtained from multiple Part A types (inpatient, home health, hospice, skilled nursing facility claims) and Part B outpatient types of Medicare Claims SAFs.

For the Fall 2019 maintenance submission, the description above still applies.

1.3. What are the dates of the data used in testing?
January 1, 2011 – December 31, 2014
January 1, 2014 – December 31, 2017

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For each year, we first included all Medicare certified facilities. The following table (Table 1) shows the count of the facilities each year, before and after exclusions were applied; we also report percent excluded for each year.

Table 1: Count of facilities per year, before and after patient-level comorbidity exclusion.

	Facility Count		
Year	**Before Exclusions**	**After Exclusions**	**Percent Excluded**
2011	5777	5774	0.05%
2012	5955	5943	0.20%
2013	6184	6170	0.23%
2014	6422	6415	0.11%

For the Fall 2019 maintenance submission, we first included all Medicare certified facilities. The following table (Table 2) shows the count of the facilities each year, before and after exclusions were applied; we also report percent excluded for each year.

Table 2: Count of facilities per year, before and after patient-level comorbidity exclusion.

	Facility Count		
Year	Before Exclusions	After Exclusions	Percent Excluded
2014	6,423	6,411	0.2%
2015	6,618	6,599	0.3%
2016	6,896	6,857	0.6%
2017	7,129	7,099	0.4%

Table 3. Number of facilities included for testing and analysis for the years 2011-2014.

Year	# of facilities	Mean Facility size (patients)
2011	5774	67.04
2012	5943	67.10
2013	6170	67.35
2014	6415	66.91

Table 4. Number of facilities included for testing and analysis for the years 2014-2017.

Year	# of facilities	Mean Facility size (patients)
2014	6,411	59.7
2015	6,599	58.1
2016	6,857	56.7
2017	7,099	55.7

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Table 5. Count of facilities, patients, and total patient years at risk.

Year	# of facilities	# of Patients	Total Patients Years at risk
2011	5774	387097	227935.62
2012	5943	398769	234847.09
2013	6170	415576	241082.06
2014	6415	429241	246710.49

Table 6. Count of facilities, patients, and total patient years at risk.

Year	# of facilities	# of Patients	Total Patients Years at risk
2014	6,411	383,001	235,885.7
2015	6,599	383,075	231,699.4
2016	6,857	388,572	234,511.4
2017	7,099	395,189	234,996.3

The following table (Table 7) shows the facility level mean number of patients, mean age; mean values for patient years at risk, mean %females, %black, %white, and %Hispanics for each of the four years.

Table 7. Facility level mean values.

Year	# Patients	Age as of end of year	Patient Yrs at Risk	%Female	%Black	%White	%Hisp
2011	67.04	63.32	39.48	45.45	32.17	62.15	14.16
2012	67.10	63.29	39.52	45.55	32.02	62.37	14.37
2013	67.35	63.38	39.07	45.16	31.83	62.46	14.39
2014	66.91	63.50	38.46	44.85	31.71	62.42	14.42

For the Fall 2019 maintenance submission, the following table (Table 8) shows the facility level mean number of patients, mean values for patient years at risk, mean age, mean %female, %black, %white, and %Hispanic for each of the four years.

Table 8. Facility level mean values.

Year	# Patients	Patient Yrs at Risk	Age as of end of year	%Female	%Black	%White	%Hispanic
2014	59.7	36.8	62.6	44.6%	31.9%	62.0%	14.6%
2015	58.1	35.1	62.8	44.3%	32.0%	61.8%	14.8%
2016	56.7	34.2	62.9	44.0%	31.9%	61.8%	14.7%
2017	55.7	33.1	63.0	43.6%	31.7%	61.8%	14.9%

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

All reliability, validity, risk adjustment analyses are done using this data set as explained in Table 1 of Section 1.5 above.

For the test of meaningful differences, please refer section 2b.5 for details, facilities with less than 10 patient years at risk are excluded from this analysis.

Table 9. Counts of facilities before and after application of the less than 10 patient years at risk exclusion, 2011-2014.

Year	# Facilities included in the testing and analysis	# Facilities with at least 10 patient years at risk	Percent excluded
2011	5774	5138	11.01%
2012	5943	5318	10.52%
2013	6170	5441	11.82%
2014	6415	5650	11.93%

For the Fall 2019 maintenance submission, All reliability, validity, and risk adjustment analyses are done using the facility counts as reported in Table 2 of Section 1.5 above.

For the test of meaningful differences, please refer to section 2b.5 for details. Facilities with less than 10 patient years at risk are excluded from this analysis.

Table 10. Counts of facilities before and after application of the less than 10 patient years at risk exclusion, 2014-2017.

Year	# Facilities included in the testing and analysis	# Facilities with at least 10 patient years at risk	Percent excluded
2014	6,411	5,586	12.9%
2015	6,599	5,671	14.1%
2016	6,857	5,887	14.1%
2017	7,099	6,075	14.4%

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from

each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

Patient level:

- Employment status 6 months prior to ESRD
- Sex
- Race
- Ethnicity
- Medicare coverage*

**Assessed at a specific time point (e.g., at a transfusion event). The final variable for Medicare coverage in model was recoded*
1. Medicare as primary and Medicaid
2. Medicare as primary and NO Medicaid
3. Medicare Secondary or Medicare HMO
4. Non-Medicare/missing

Data on patient level SDS/SES factors obtained from Medicare claims and administrative data.

ZIP code level – Area Deprivation Index (ADI) elements from Census data:

- Unemployment rate (%)
- Median family income (rescaled as (income-60,000)/10,000)
- Income disparity
- Families below the poverty level (%)
- Single-parent households w/ children <18 (%)
- Home ownership rate (%)
- Median home value (rescaled as (homevalue-200,000)/100,000)
- Median monthly mortgage (rescaled as (mortgage-1,500)/1,000)
- Median gross rent (rescaled as (rent-900)/1,000)
- Population (aged 25+) with <9 years of education (%)
- Population (aged 25+) w/o HS diploma (%)

For the Fall 2019 maintenance submission,

Patient level:

- Employment status 6 months prior to ESRD
- Sex
- Race
- Ethnicity
- Medicare coverage*

**Assessed at a specific time point (e.g., at a transfusion event). The final variable for Medicare coverage in the model was recoded as:*
1. Medicare as primary and Medicare Secondary
2. Medicare as primary and Medicaid (dual eligible)

3. Non-Medicare/missing

ZIP code level – Area Deprivation Index (ADI) from Census data (2015). Based on patient zip-code.

_____

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

The reliability of the STrR was assessed using data among ESRD dialysis patients during 2011-2014. If the measure were a simple average across individuals in the facility, the usual approach for determining measure reliability would be a one-way analysis of variance (ANOVA), in which the between and within facility variation in the measure is determined. The inter-unit reliability (IUR) measures the proportion of the measure variability that is attributable to the between-facility variance. The STrR, however, is not a simple average and we instead estimate the IUR using a bootstrap approach, which uses a resampling scheme to estimate the within facility variation that cannot be directly estimated by ANOVA. A small IUR (near 0) reveals that most of the variation of the measures between facilities is driven by random noise, indicating the measure would not be a good characterization of the differences among facilities, whereas a large IUR (near 1) indicates that most of the variation between facilities is due to the real difference between facilities.

Here we describe our approach to calculating IUR. Let $T_1,\ldots,T_N$ be the STrR for these facilities. Within each facility, select at random and with replacement B bootstrap samples. Our numerical experiments reveal that B=100 is sufficient. That is, if the ith facility has n_i subjects, randomly draw with replacement n_i subjects from those in the same facility, find their corresponding $STrR_i$ and repeat the process B (say, 100) times. Thus, for the ith facility, we have bootstrapped STrRs of $T^*_1,\ldots,T^*_{100}$ …, . Let S_i be the sample variance of this bootstrap sample. From this it can be seen that

$$s^2_{t,w} = \frac{\sum_{i=1}^{N}\left[(n_i-1)S_i^2\right]}{\sum_{i=1}^{N}(n_i-1)},$$

is a bootstrap estimate of the within-facility variance in the STtR, namely $\sigma^2_{t,w}$. Calling on formulas from the one way analysis of variance, an estimate of the overall variance of T_i is

$$s^2_t = \frac{1}{n'(N-1)}\sum_{i=1}^{N} n_i(T_i-\bar{T})^2,$$

where

$$\bar{T} = \Sigma\, n_i\, T_i / \Sigma\, n_i$$

is the weighted mean of the observed STrR and

$$n' = \frac{1}{N-1}\left(\sum n_i - \sum n_i^2 \Big/ \sum n_i\right)$$

is approximately the average facility size (number of patients per facility). Note that s_t^2 is an estimate of $\sigma_b^2 + \sigma_{t,w}^2$, where σ_b^2 is the between-facility variance, the true signal reflecting the differences across facilities. Thus, the IUR = $\sigma_b^2/(\sigma_b^2 + \sigma_{t,w}^2)$, can be estimated by $(s_t^2 - s_{t,w}^2)/s_t^2$.

The STrR calculation only included facilities with at least 10 patient years at risk.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

The STrR calculation only included facilities with at least 10 patient years at risk. Overall, we found that IURs for the one-year STrR have a range of 0.60-0.66 across the years 2011, 2012, 2013 and 2014, which indicates that around two-thirds of the variation in the one-year STrR can be attributed to the between-facility differences and one-third to within-facility variation. This value of IUR indicates a **moderate degree of reliability**. When stratified by facility size, we find that, as expected, larger facilities have greater IUR.

Table 11: IUR for One-year STrR, Overall and by Facility Size, 2011-2014.

	2011		**2012**		**2013**		**2014**	
Facility Size	IUR	N	IUR	N	IUR	N	IUR	N
all	0.64	5142	0.66	5319	0.65	5442	0.60	5651

	2011		2012		2013		2014	
Small (<=46)	0.41	1714	0.41	1828	0.39	1840	0.30	1934
Medium (47–78)	0.55	1699	0.56	1753	0.55	1823	0.50	1941
Large (>=79)	0.78	1729	0.79	1738	0.79	1779	0.78	1776

For the Fall 2019 maintenance submission, we found that IURs for the one-year STrR have a range of 0.63-0.68 across the years 2014, 2015, 2016 and 2017, which indicates that around two-thirds of the variation in the one-year STrR can be attributed to the between-facility differences and one-third to within-facility variation. This value of IUR indicates a **moderate degree of reliability**.

Table 12: IUR for One-year STrR, 2014-2017.

	2014		2015		2016		2017	
	IUR	**N**	**IUR**	**N**	**IUR**	**N**	**IUR**	**N**
All	0.66	5,587	0.68	5,672	0.64	5,888	0.63	6,076

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This value of IUR indicates a moderate degree of reliability. When stratified by facility size, we find that, as expected, larger facilities have greater IUR.

For the Fall 2019 maintenance submission, the interpretation is the same; the value of IUR indicates a moderate degree of reliability. When stratified by facility size, we find that, as expected, larger facilities have greater IUR.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Validity was assessed using Poisson regression models to measure the association between facility level the 2014 Standardized Mortality Ratio (SMR, NQF 0369) and 2014 Standardized Hospitalization Ratio (SHR, NQF 1463) and tertiles of STrR. Facility-level STrR were divided into tertiles (T1 to T3) and the relative risk (RR) of mortality (and hospitalization, separately) was calculated for each tertile, using the T1 as the reference group. Thus, a RR>1.0 would indicate a higher relative risk of mortality or hospitalization, compared to the highest performance tertile (T1) of STrR.

Validity was also assessed using a Poisson regression model to measure the association between facility level STrR and tertiles of % of patients with Hgb < 10. Facility-level % of patients with Hgb < 10 were divided into tertiles (T1 to T3) and relative risk (RR) of transfusions were calculated for each tertile, using the T1 as the reference group. Thus, a RR>1.0 would indicate a higher relative risk of transfusion, compared to the highest performance tertile(T1) of % of patients with Hgb < 10.

For the Fall 2019 maintenance submission, validity was again assessed using Poisson regression models to measure the association between the facility level 2017 Standardized Mortality Ratio (SMR, NQF 0369) and 2017 Standardized Hospitalization Ratio (SHR, NQF 1463) and tertiles of STrR. Facility-level STrR were divided into tertiles (T1 to T3) and the relative risk (RR) of mortality (and hospitalization, separately) was calculated for each tertile, using T1 as the reference group. Thus, a RR>1.0 would indicate a higher relative risk of mortality or hospitalization, compared to the highest performance tertile (T1) of STrR. We expect the risk of mortality and hospitalization, respectively, will be positively associated with higher tertiles of STrR.

Validity was also assessed using a Poisson regression model to measure the association between facility level STrR and tertiles of % of patients with Hgb < 10. Facility-level % of patients with Hgb < 10 were divided into tertiles (T1 to T3) and relative risk (RR) of transfusions were calculated for each tertile, using T1 as the reference group. Thus, a RR>1.0 would indicate a higher relative risk of transfusion, compared to the highest performance tertile (T1) of % of patients with Hgb < 10. We expect the risk of a transfusion event will be positively associated with higher tertiles of % of patients with Hgb < 10.

In May 2012 there was an assessment of the measure's face validity based on polling of a CMS Technical Expert Panel (TEP). This assessment is carried forward to the Fall 2019 maintenance submission.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Association of STrR with other facility-level outcomes

Tertiles of STrR were defined as follows:

T1: 0-<0.66
T2: 0.66-<1.15
T3: 1.15-<5.66
*T1 as Reference

Results from the Poisson model indicated that the STrR tertiles were significantly associated with both SMR and SHR. For the 2014 SMR, the relative risk of mortality increased as the STrR tertiles increased from the reference group (tertile 1). For tertile 2, RR=1.06 (95% CI: 1.04, 1.08; p<0.001), and for tertile 3, RR=1.14 (95% CI: 1.12, 1.16; p<0.001). Similarly for 2014 SHR, the relative risk of hospitalization increased as the STrR tertiles increased from the reference group (tertile 1) with the lowest risk in tertile 1. For tertile 2, RR=1.11 (95% CI: 1.10, 1.11; p<0.001), and for tertile 3, RR=1.29 (95% CI: 1.29, 1.30; p<0.001).

For the Fall 2019 maintenance submission, tertiles of STrR were defined as follows:
T1: 0-<0.70
T2: 0.70-<1.13
T3: 1.13-<7.1
*T1 as Reference

Results from the Poisson models indicated that the STrR was significantly associated with the risks of mortality and hospitalization. For the 2017 SMR, the relative risk of mortality increased as the STrR tertiles increased from the reference group (tertile 1). For tertile 2, RR=1.09 (95% CI: 1.07, 1.11; p<0.001), and for tertile 3, RR=1.17 (95% CI: 1.15, 1.19; p<0.001). Similarly for 2017 SHR, the relative risk of hospitalization increased as the STrR tertiles increased from the reference group (tertile 1) with the lowest risk in tertile 1. For tertile 2, RR=1.15 (95% CI: 1.15, 1.16; p<0.001), and for tertile 3, RR=1.32 (95% CI: 1.32, 1.32; p<0.001).

Association of STrR with facility-level intermediate anemia management outcome

Tertiles of % of patients with Hgb < 10 were defined as follows:

T1: 0-<9.5%
T2: 9.5%-<16.5%
T3: 16.5%-<85.3%
*T1 as Reference

Results from the Poisson model indicated that the % of patients with Hgb < 10 was significantly associated with the risks of transfusion. The relative risk of transfusions increased as the tertiles of % of patients with Hgb < 10 increased from the reference group (tertile 1). For tertile 2, RR=1.15 (95% CI: 1.13, 1.18; p<0.001), and for tertile 3, RR=1.31 (95% CI: 1.28, 1.33; p<0.001).

For the Fall 2019 maintenance submission, tertiles of % of patients with Hgb < 10 were defined as follows:
T1: 3.7-<17.5%
T2: 17.5%-<22.3%
T3: 22.3%-<55.4%
*T1 as Reference

Results from the Poisson model indicated that the % of patients with Hgb < 10 was significantly associated with the risks of transfusion. The relative risk of transfusions increased as the tertiles of % of patients with Hgb < 10 increased from the reference group (tertile 1). For tertile 2, RR=1.17 (95% CI: 1.15, 1.20; p<0.001), and for tertile 3, RR=1.44 (95% CI: 1.42, 1.47; p<0.001).

Results of TEP Vote Establishing Face Validity of Standardized Transfusion Ratio

Six out of six voting members of CMS's 2012 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 and in some cases, increased risk of ESA-associated adverse events, such as hereditary anemia, chronic bone marrow failure conditions and active cancer. This assessment is carried forward to the Fall 2019 maintenance submission.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)
The overall measure demonstrates face validity based on the structured 2012 TEP vote.

Furthermore, testing of the measure supports construct validity. The positive correlation between this measure and SMR and SHR respectively indicates that facilities with more transfusions than would be expected based on national rates, also have higher standardized mortality and standardized hospitalization rates.

In addition to the demonstrated association between STrR and other facility outcomes, the above results demonstrate the association between facility-level achieved hemoglobin, an intermediate outcome reflecting facility anemia management processes, and STrR. The results of dialysis facility

achieved hemoglobins, grouped into tertiles, demonstrates statistically significant differences across tertiles with reassuring stepwise increments of STrR between tertiles, suggesting “dose effect”.

For the Fall 2019 maintenance submission, the measure is maintained on the basis of face validity. The testing results continue to support the measure’s construct validity. The positive correlation between STrR with SMR and SHR, respectively, indicates that facilities with more transfusions than would be expected based on national rates, also have higher standardized mortality and standardized hospitalization rates.

In addition to the demonstrated association between STrR and other facility outcomes, the above results demonstrate the association between facility-level achieved hemoglobin, an intermediate outcome reflecting facility anemia management processes, and STrR. The results of dialysis facility achieved hemoglobins, grouped into tertiles, demonstrates statistically significant differences across tertiles with reassuring stepwise increments of STrR between tertiles, suggesting “dose effect”.

_____

2b2. EXCLUSIONS ANALYSIS

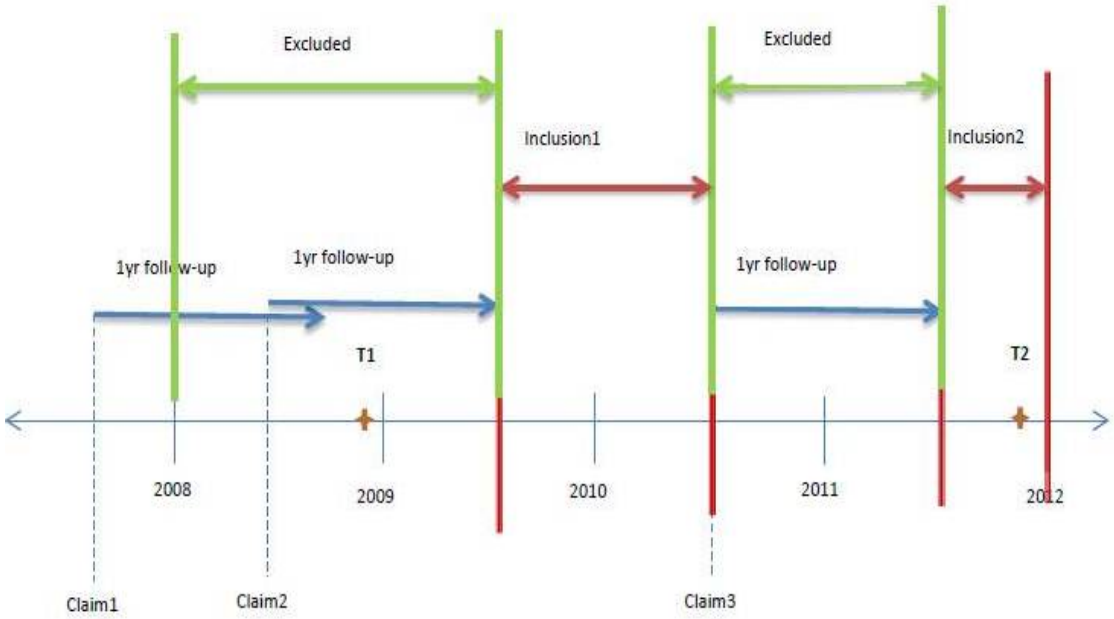
NA ☐ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Transfusions associated with transplant hospitalization are excluded as they mark a transition of care from the dialysis facility to a transplant team. This convention is used with other dialysis facility measures developed and previously endorsed by NQF (like SHR NQF #1463 http://www.qualityforum.org/QPS/1463) and SMR NQF #0369 http://www.qualityforum.org/QPS/0369)

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 free of claims that contain diagnoses on this exclusion list. We assessed the predictive power of these 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 event as the dependent variable.

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 Appendix) 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 the figure). 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 the presence of the 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.

For the Fall 2019 maintenance submission, the information above on the approach to exclusions remains current.

For analyses related to the exclusion of Medicare Advantage patients, see section below on Missing Data.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Multivariate logistic regression with transfusion event 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 event was coded as a binary variable (1 if transfusion). Results using 2011 data showed that a 1-year look back period for each of the exclusion comorbidities was a significant predictor of RBC transfusion events with odds ratio ranging from 1.2 to 3.2.

The following tables show percent of patient years at risk and number of patients excluded as a result of the above mentioned exclusion strategy.

Table 13: Percent of patient years at risk (PYR) excluded each year.

	Patient years at risk		
Year	**Before Exclusions**	**After Exclusions**	**Percent Excluded**
2011	287056.42	227935.62	20.60%
2012	296411.19	234847.09	20.77%
2013	302026.41	241082.06	20.18%
2014	308375.2	246710.49	20.00%

For the Fall 2019 maintenance submission, the following tables show percent of patient years at risk and number of patients excluded as a result of the above mentioned exclusion strategy.

Table 14: Percent of patient years at risk (PYR) excluded each year.

	Patient Years at Risk		
Year	Before Exclusions	After Exclusions	Percent Excluded
2014	353,922.4	235,885.7	33.4%
2015	363,530.8	231,699.4	36.3%
2016	375,529.6	234,511.4	37.6%
2017	382,669.3	234,996.3	38.6%

Table 15: Number of patients and percent excluded each year.

	Number of Patients		
Year	**Before Exclusions**	**After Exclusions**	**Percent Excluded**
2011	452134	387097	14.38%
2012	468592	398769	14.90%
2013	486644	415576	14.60%
2014	503016	429241	14.67%

Table 16: Number of patients and percent excluded each year.

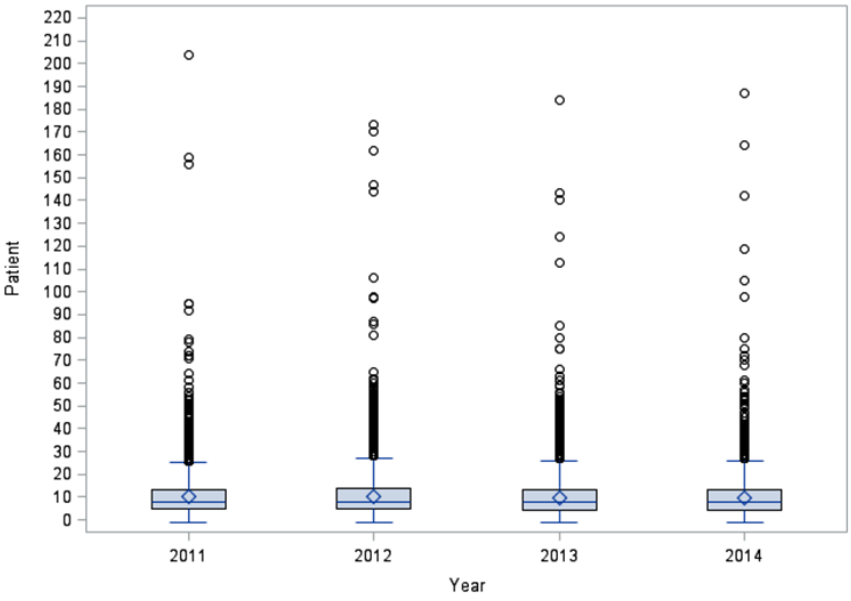
	Number of Patients		
Year	Before Exclusions	After Exclusions	Percent Excluded
2014	533,626	383,001	28.2%
2015	549,823	383,075	30.3%
2016	568,117	388,572	31.6%
2017	587,906	395,189	32.8%

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The list of comorbidities described in section 2b3.1 have been associated with ESA resistance and higher risk of transfusion, as well as increased risk of ESA use. Based on these factors, they require different anemia management practices that this measure is not intended to address; hence the need for the comorbidity exclusions. The Technical Expert Panel had also recommended these exclusions. As described in Section 2b3.2 patients with exclusion comorbidities are at a higher risk to get transfused than patients that do not have these comorbidities.

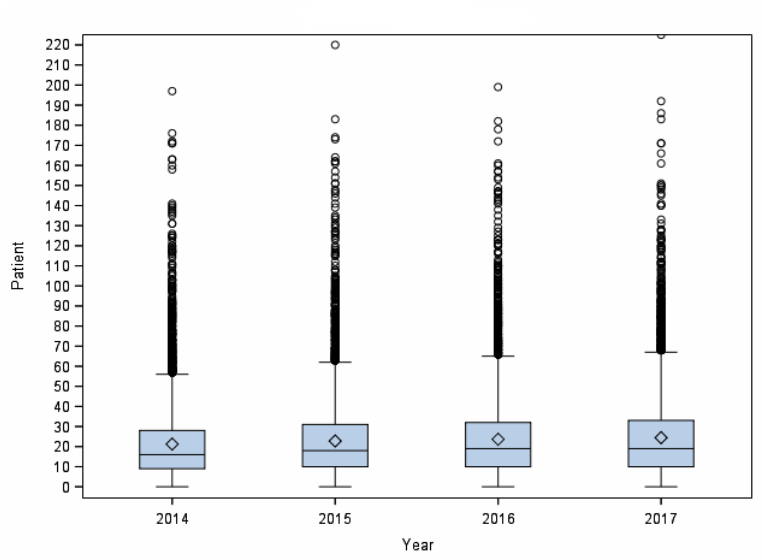
We also checked the distribution of patients excluded at the facility level and the boxplot shows that there is variability in the number of patients excluded among facilities. The numbers of patients with the exclusion comorbidities are not uniformly distributed across facilities thereby demonstrating the need for an exclusion strategy.

Figure 2: Distribution of Excluded Patients at facility level for 2011-2014



For the Fall 2019 maintenance submission, the interpretation remains consistent with what was provided in the 2016 submission.

Figure 3: Distribution of Excluded Patients at facility level for 2014-2017



2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 40 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

The following description of the model provided in 2016 applies to the fall 2019 maintenance submission.

The denominator of the STrR stems from a proportional rates model (Lawless and Nadeau, 1995; Lin et al., 2000; Kalbfleisch and Prentice, 2002). This is the recurrent event analog of the well-known proportional hazards or Cox model (Cox, 1972; Kalbfleisch and Prentice, 2002). To accommodate large-scale data, we adopt a model with piecewise constant baseline rates (e.g. Cook and Lawless, 2007) and the computational methodology developed in Liu, Schaubel and Kalbfleisch (2012). The modeling process has two stages. At stage I, a stratified model is fitted to the national data with piecewise-constant baseline rates and stratification by facility. Specifically, the model is of the following form:

$$Pr(\text{transfusion on day } t \text{ given covariates } X) = r_{0k}(t)\exp(\beta' X_{ik})$$

where X_{ik} is the vector of covariates for the (i,k)th patient and β is the vector of regression coefficients. The baseline rate function $r_{0k}(t)$ is assumed specific to the k^{th} facility, which is assumed to be a step function with break points at 6 months, 1 year, 2 years, 3 years and 5 years since the onset of dialysis. 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 stratification on facilities is important in this phase to avoid bias due to possible confounding between covariates and facility effects.

The patient characteristics X_{ik} included in the stage I model 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), 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), nursing home status, BMI at incidence, individual comorbidities at incidence reported on the Medical Evidence Form (CMS-2728), calendar year, and two-way interaction terms between age and duration and cause of ESRD. Nursing home status is identified as in or not in a nursing home in the previous calendar year. BMI is included as a log-linear term. Categorical indicator variables are included as covariates in the stage I model to flag records missing values for cause of ESRD, 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 two categorical indicator variables are included to flag records with having no comorbidities and having at least one comorbidity at incidence reported on the Medical Evidence Form. These variables have a value of 1 if the patient is having no comorbidities or having at least one comorbidity and a value of 0 otherwise.

At stage II, the relative risk estimates from the first stage are used to create offsets and an unstratified model is fitted to obtain estimates of an overall baseline rate function. That is, we estimate a common baseline rate of transfusions, r0(t), across all facilities by considering the model

$$Pr(\text{transfusion on day } t \text{ given covariates } X) = r_0(t)\, R_{ik,}$$

where $R_{ik} = \exp(\beta' X_{ik})$ is the estimated relative risk for patient i in facility k estimated from the stage I. In our computation, we assume the baseline to be a step function with 6 unknown parameters, $\alpha_1, \ldots, \alpha_6$, to estimate. These estimates are used to compute the expected number of transfusions given a patient's characteristics.

Specifically, let t_{iks} represent the number of days that patient i from facility k is under observation in the sth time interval with estimated rate α_s. The corresponding expected number of transfusions in the sth interval for this patient is calculated as:

$$E_{iks} = \alpha_s\, t_{iks}\, R_{ik} .$$

It should be noted that tiks and hence E_{iks} can be 0 if patient i from facility k is never at risk during the sth time interval. Summing the E_{iks} over all 6 intervals and all N_k patients in a given facility, k, gives

$$Exp = \sum_{i=1}^{N} \sum_{s=1}^{6} E_{iks} = \sum_{i=1}^{N} \sum_{s=1}^{6} \alpha_s\, t_{iks}\, R_{ik}$$

The patient characteristics included in the stage 1 model as covariates are:

- Age: We determine each patient's age for the birth date provided in the SIMS and REMIS databases and group patients into the following categories: 0-14 years old, 15-24 years old, 25-44 years old, 45-59 years old, 60-74 years old, or 75+ years old.
- Diabetes as cause of ESRD: We determine each patient's primary cause of ESRD from his/her CMS-2728.
- Duration of ESRD: We determine each patient's length of time on dialysis using the first service date from his/her CMS-2728, claims history (all claim types), the SIMS database and the SRTR database and categorize as 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.
- Nursing home status: Using the Nursing Home Minimum Dataset, we determine if a patient was in a nursing home the previous year.
- BMI at incidence: We calculate each patient's BMI as the height and weight provided on his/her CMS 2728. BMI is included as a log-linear term.
- Comorbidities at ESRD incidence are determined using a selection of comorbidities reported on the CMS-2728 namely, alcohol dependence, atherosclerotic heart disease, cerebrovascular disease, chronic obstructive pulmonary disease, congestive heart failure, diabetes (includes currently on insulin, on oral medications, without medications, and diabetic retinopathy), drug dependence, inability to ambulate, inability to transfer, malignant neoplasm, cancer, other

cardiac disease, peripheral vascular disease, and tobacco use (current smoker). Each comorbidity is included as a separate covariate in the model.

- Calendar year
- Categorical indicator variables are included as covariates in the stage I model to account for records with missing values for cause of ESRD, comorbidities at incidence (missing CMS-2728), and BMI. These variables have a value of 1 if the patient is missing the corresponding variable and a value of 0 otherwise. Another categorical indicator variable is included as a covariate in the stage 1 model to flag records where the patient has at least one of the incident comorbidities listed earlier. This variable has a value of 1 if the patient has at least one of the comorbidities and a value of 0 otherwise.

Beside main effects, two-way interaction terms between age and duration and cause of ESRD are also included:

- Diabetes as cause of ESRD*Duration of ESRD
- Diabetes as cause of ESRD*Age

The same coefficient weights are used as in the Standardized Hospitalization Ratio (see www. dialysisdata.org; NQF #1463 http://www.qualityforum.org/QPS/1463).

References:

Cox, D.R. (1972) Regression Models and Life Tables (with Discussion). J. Royal statistical Society, Series B, 34, 187-220.

Cook, R. and Lawless, J. The Statistical Analysis of Recurrent Events. New York: Springer. 2007.

Cook, R. and Lawless, J. Marginal analysis of recurrent events and a terminal event. Statistics in Medicine 1997; 16: 911-924.

Kalbfleisch, J.D. and Prentice, R. L. The Statistical Analysis of Failure Time Data. Wiley, New York, 2002.

Lawless, J. F. and Nadeau, C. Some simple and robust methods for the analysis of recurrent events, Technometrics, 37 1995, 355-364.

Lin, D.Y., Wei, L.J., Yang, I. and Ying, Z. Semi parametric regression for the mean and rate functions of recurrent events, Journal of the Royal Statistical Society Series B, 62, 2000, 771-730

Liu, D., Schaubel, D.E. and Kalbfleisch, J.D. Computationally efficient marginal models for clustered recurrent event data, University of Michigan Department of Biostatistics Technical Reports, 2010.

2b3.2. If an outcome or resource use component measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

In the table below, we list results from the Stage 1 model described above that includes the selected patient characteristics and other risk adjustors. For a given covariate, the parameter estimate represents the log of the rate ratio (recurrent event version of the relative risk). All covariates have face validity from a clinical perspective. We assume these selected covariates do not reflect the quality of facility care, nor, disparities in care. With the exceptions of BMI=missing and cancer, all main effects are statistically significant at 0.05 level.

For the Fall 2019 maintenance submission, the description of the model and adjustments remains the same as described in the 2016 submission. With the exception of BMI=missing, all main effects are statistically significant at 0.05 level.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

- ☒ **Published literature**
- ☒ **Internal data analysis**
- ☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

Table 17. Model Coefficients for STrR, 2014

Covariate	Coefficient	P-value
Cause of ESRD		
Diabetes	-0.118	<.0001
Missing	0.188	<.0001
Age		
18-24	0.084	<.0001
25-44	-0.196	<.0001
45-59	-0.18	<.0001
60-74	Reference	
75+	0.035	<.0001
BMI		
Log BMI	-0.247	<.0001
BMI missing	0.024	0.19
Calendar year		
2011	Reference	
2012	0.068	<.0001

Covariate	Coefficient	P-value
2013	0.027	<.0001
2014	-0.08	<.0001
In nursing home the previous year	0.489	<.0001
Diabetes as cause of ESRD & time on ESRD interaction term		
91 days-6 months	Reference	
6 months-1 year	0.068	0.001
1-2 years	0.128	<.0001
2-3 years	0.135	<.0001
3-5 years	0.09	<.0001
5+ years	0.044	0.014
Age & diabetes as cause of ESRD interaction term		
0-14		
15-24	0.166	0.09
25-44	0.228	<.0001
45-59	0.098	<.0001
60-74	Reference	
75+	0.008	0.445
Incident comorbidities		
atherosclerotic heart disease	0.071	<.0001
other cardiac disease	0.065	<.0001
congestive heart failure	0.049	<.0001
inability to ambulate	0.108	<.0001
chronic obstructive pulmonary disease	0.168	<.0001
inability to transfer	0.097	<.0001
cancer	0.008	0.541
diabetes	0.085	<.0001
peripheral vascular disease	0.134	<.0001
cerebrovascular disease	0.02	0.005
tobacco use (current smoker)	0.135	<.0001
alcohol dependence	0.117	<.0001
drug dependence	0.097	<.0001
At least one incident comorbidity	0.088	<.0001
Incident comorbidity missing	0.068	0.008

Table 18. Model Coefficients for STrR, 2017

Covariate	Coefficient	P-value
Cause of ESRD		
Diabetes	-0.128	<.0001
Missing	0.065	0.023
Age		
18-24	0.210	<.0001
25-44	-0.084	<.0001
45-59	-0.117	<.0001
60-74	Reference	
75+	-0.054	<.0001
BMI		
Log BMI	-0.222	<.0001
BMI missing	0.010	0.604
Calendar year		
2014	Reference	
2015	-0.062	<.0001
2016	-0.108	<.0001
2017	-0.162	<.0001
In nursing home the previous year	0.685	<.0001
Diabetes as cause of ESRD & time on ESRD interaction term		
91 days-6 months	Reference	
6 months-1 year	0.070	0.001
1-2 years	0.147	<.0001
2-3 years	0.135	<.0001
3-5 years	0.108	<.0001
5+ years	0.067	0.000
Age & diabetes as cause of ESRD interaction term		
0-14		
15-24	0.498	<.0001
25-44	0.211	<.0001
45-59	0.108	<.0001
60-74	Reference	
75+	0.041	<.0001
Incident comorbidities		
atherosclerotic heart disease	0.065	<.0001
other cardiac disease	0.091	<.0001

Covariate	Coefficient	P-value
congestive heart failure	0.038	<.0001
inability to ambulate	0.079	<.0001
chronic obstructive pulmonary disease	0.176	<.0001
inability to transfer	0.073	<.0001
cancer	-0.059	<.0001
diabetes	0.078	<.0001
peripheral vascular disease	0.143	<.0001
cerebrovascular disease	-0.030	<.0001
tobacco use (current smoker)	0.160	<.0001
alcohol dependence	0.105	<.0001
drug dependence	0.109	<.0001
At least one incident comorbidity	0.094	<.0001
Incident comorbidity missing	0.092	0.001

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

The table below shows the parameter estimates for patient and area level SDS/SES variables tested based on a model that included these variables along with the original covariates.

Table 19. Parameter estimates for patient and area level SDS/SES variables

Covariate	Estimates	P-value	Hazard Ratio
Sex: Female	0.163	<.0001	1.177
Race			
White	ref		
Black	-0.048	<.0001	0.953
Asian/Pacific Islander	-0.180	<.0001	0.835
Native American	-0.044	0.058	0.957
Other	-0.031	0.114	0.970
Hispanic	-0.174	<.0001	0.840
Employment status			
Employed	ref		
Unemployed	0.119	<.0001	1.126
Other	0.145	<.0001	1.156
Medicare coverage			

Covariate	Estimates	P-value	Hazard Ratio
Medicare as primary w/o Medicaid	ref		
Medicare as primary with Medicaid	0.025	<.0001	1.025
Medicare as secondary /Medicare HMO	0.724	<.0001	2.062
Non-Medicare/missing	-0.025	0.585	0.975
ADI			
Unemployment rate (%)	0.000	0.829	1.000
Median family income	-0.002	0.502	0.998
Families below the poverty level (%)	0.000	0.868	1.000
Single-parent households w/ children <18 (%)	-0.001	0.176	0.999
Home ownership rate (%)	0.001	0.015	1.001
Median home value	0.011	0.019	1.011
Median monthly mortgage	-0.003	0.826	0.997
Median gross rent	0.007	0.680	1.007
Population (aged 25+) w/o HS diploma (%)	-0.001	0.275	0.999
Income disparity	0.015	0.009	1.016

Patient-level SDS/SES: Compared to males, females were more likely to receive transfusions (HR=1.17; $p<0.01$). Compared to white patients, black patients were less likely to receive transfusions (HR=0.95, $p<0.01$). Hispanics were less likely to have transfusions (HR=0.84; $p<0.01$), compared to non-Hispanics. Compared to Medicare only patients, patients with both Medicare/ Medicaid (HR=1.03, $p<0.01$) and Medicare as secondary /Medicare HMO (HR=2.06, $p<0.01$) were more likely to have transfusions. As for employment status, unemployed and "other" patients were more likely to have transfusions (HR=1.13; $p<0.01$; HR=1.16; $p<0.01$), compared to employed patients. Note that for employment categories, the "Other" category represents diverse patient groups with regards to SES, such as students, homemakers, and those who are retired.

Area-level SDS/SES: Area-level effects were generally all very small and most not statistically significant, with the exception of home ownership rate, median home value, and income disparity.

For the Fall 2019 maintenance submission, the table below shows the parameter estimates for patient and area level SDS/SES variables tested based on a model that included these variables along with the original covariates.

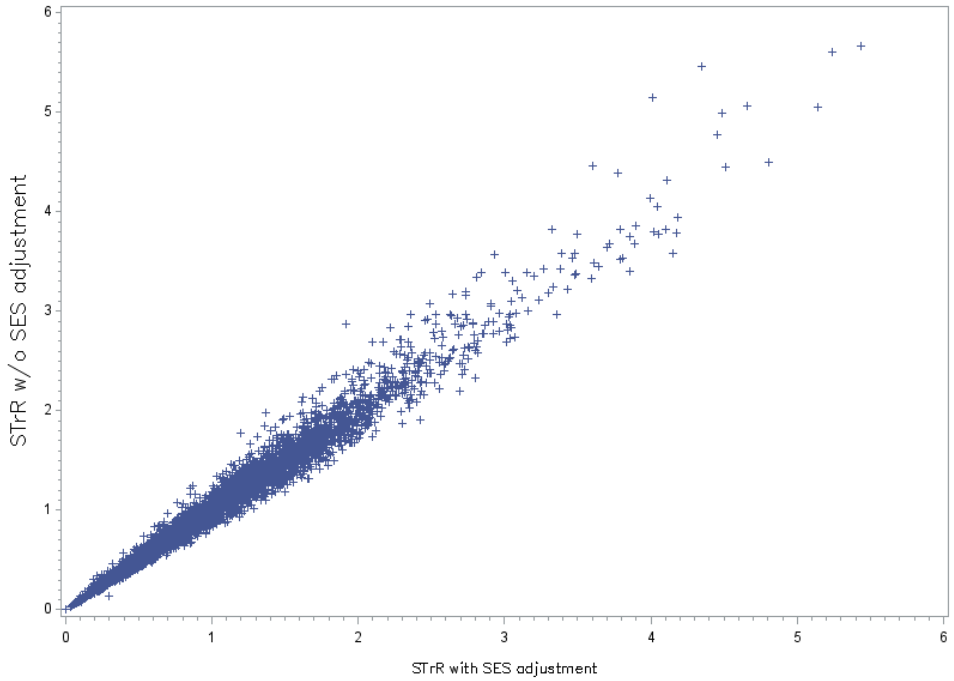
Table 20. Parameter estimates for patient and area level SDS/SES variables

Covariate	Estimates	P-value	Hazard Ratio
Sex: Female	0.151	<.0001	1.163
Race			
White	ref		
Black	-0.063	<.0001	0.939
Asian/Pacific Islander	-0.164	<.0001	0.849
Native American	-0.043	0.036	0.958
Other	0.009	0.814	1.009
Hispanic	-0.180	<.0001	0.835
Employment status			
Employed	ref		
Unemployed	0.093	<.0001	1.098
Other	0.105	<.0001	1.111
Medicare coverage			
Medicare as primary and secondary	ref		
Medicare as primary with Medicaid	-0.044	<.0001	0.957
Non-Medicare/missing	-0.232	<.0001	0.793
ADI	0.001	<.0001	1.001

Patient-level SES: Compared to males, female sex was associated with a higher risk of transfusions (HR=1.16; $p<0.01$). Hispanic ethnicity was associated with a lower risk of transfusions (HR=0.83; $p<0.01$), compared to non-Hispanics. Compared to Medicare primary and secondary patients, having Medicare and Medicaid only and no Medicare/missing was associated with a lower risk of transfusions (HR=0.95, $p<0.01$; HR=0.79, $p<0.01$). Unemployed and patients with "other or unknown" employment was associated with a higher risk of transfusions (HR=1.10; $p<0.01$; HR=1.11; $p<0.01$), compared to employed patients.

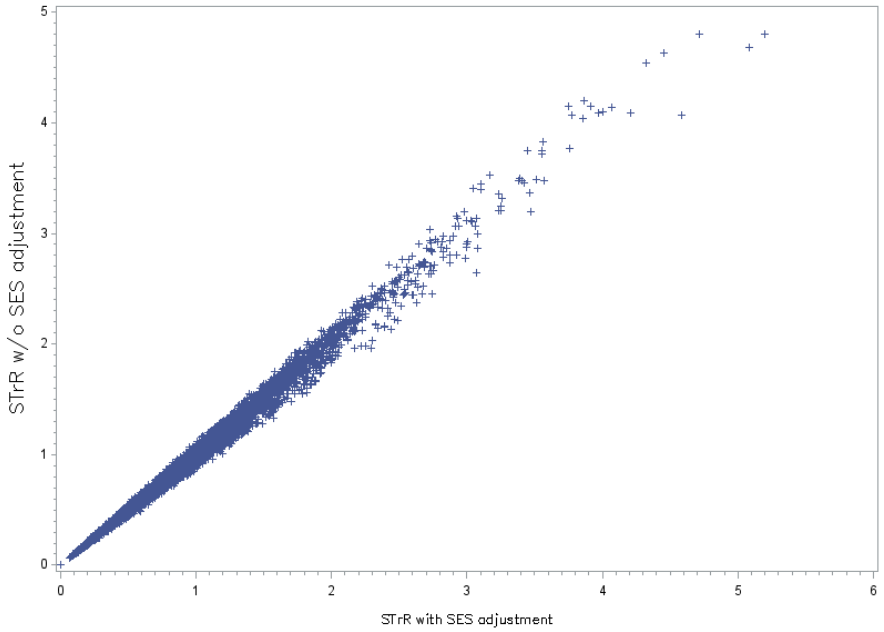
Area-level SES: The impact of area-level SES as measured by the ADI showed very little difference in association with risk of transfusions compared to different levels of area-deprivation (HR=1.001; $p<0.01$).

Correlation between STrRs with and without SDS/SES adjustment in 2014:



The standard and SDS/SES-adjusted STrR were highly correlated at 0.99 (p<.001).

For the Fall 2019 maintenance submission, correlation between STrRs with and without SDS/SES adjustment (using 2017 data):



*For readability purposes, the graph excludes one extreme outlier facility that was included in the calculation.

For the Fall 2019 maintenance submission, the standard and SDS/SES-adjusted STrR were highly correlated at 0.99 (p<.001).

Table 21. Facility performance on STrR, with and without adjustment for SDS/SES factors

STrR with SDS/SES	STrR w/o SDS/SES			
	Worse than expected	As expected	Better than expected	Total
Worse than expected	315	31	0	346(6.1%)
As expected	51	5225	6	5282(93.5%)
Better than expected	0	3	19	22(0.4%)
Total	366(6.5%)	5259(93.1%)	25(0.4%)	5650

After adjustment for SDS/SES, 91 facilities (1.6%) changed performance categories. 54 were upgraded (3 from as expected to better; 51 from worse to as expected) and 37 were degraded (6 from better to as expected; 31 from as expected to worse).

Table 22. Facility performance on STrR, with and without adjustment for SDS/SES factors

	STrR w/o SDS/SES			
STrR with SDS/SES	Worse than expected	As expected	Better than expected	Total
Worse than expected	345	16	0	361(5.9%)
As expected	29	5668	4	5701(93.8%)
Better than expected	0	3	10	13(0.2%)
Total	374(6.2%)	5687(93.6%)	14(0.2%)	6075

After adjustment for SDS/SES, 52 facilities (0.9%) changed performance categories. 32 were upgraded (3 from as expected to better; 29 from worse to as expected) and 20 were degraded (4 from better to as expected; 16 from as expected to worse).

Sex and several SDS/SES factors predict transfusion events in the patient-level model. However, inclusion of the complete set of patient sociodemographic variables, including sex, insurance status and

employment status, and the area-level indicators, shifts facility performance ranking for only a small fraction of dialysis facilities. Given the relatively constant distribution of sexes in US dialysis facilities, this demographic variable has little effect on dialysis facility-level transfusion event rates. Regarding employment and insurance status, we believe the association between transfusion events and these factors represent disparities in access to medical care and, therefore we do not believe that they are appropriate risk adjustors for a quality measure. Similarly, among the area-level indicators, all are assumed to reflect levels of economic disadvantage that represent differential access to care. For this reason we decided it was not appropriate to adjust for these differences.

This interpretation continues to apply to the Fall 2019 maintenance submission.

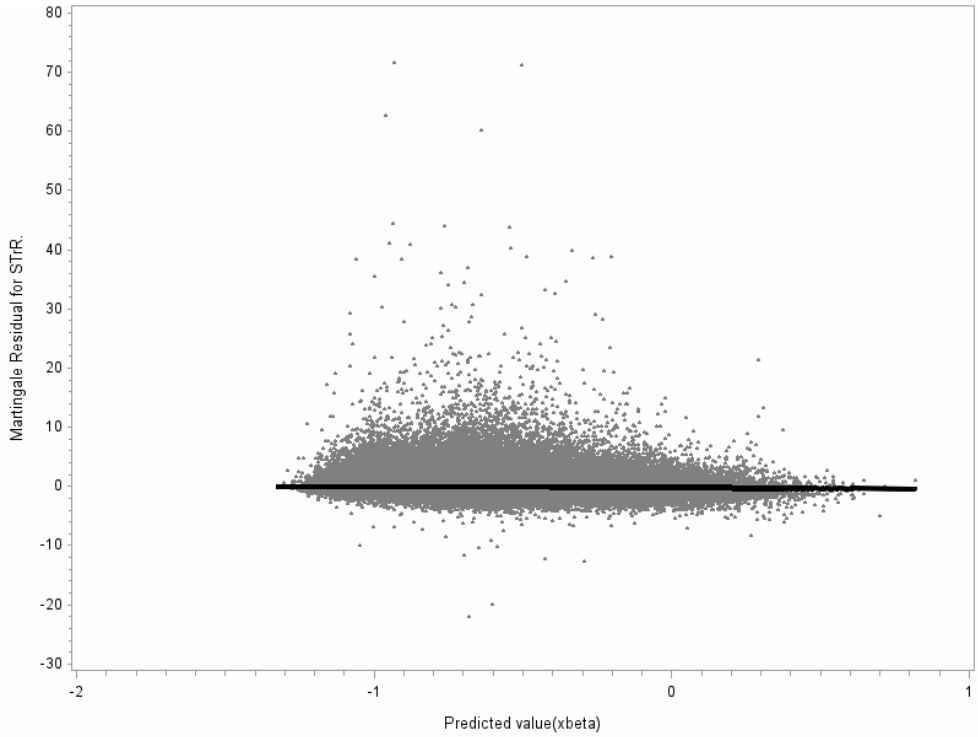
2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

Martingale residuals (Barlow and Prentice, 1988) are an important tool for checking the fit of a Cox regression model or, a model analogous to a Cox model; including the one we fitted at Stage 1. Martingale residual plots are used to investigate the lack of fit of a model. We examined the residual plot and it did not indicate problems with model fit. The LOESS curve of martingale residuals by predicted value (Figure 3) shows that the mean of the residuals is flat indicating no lack of fit.

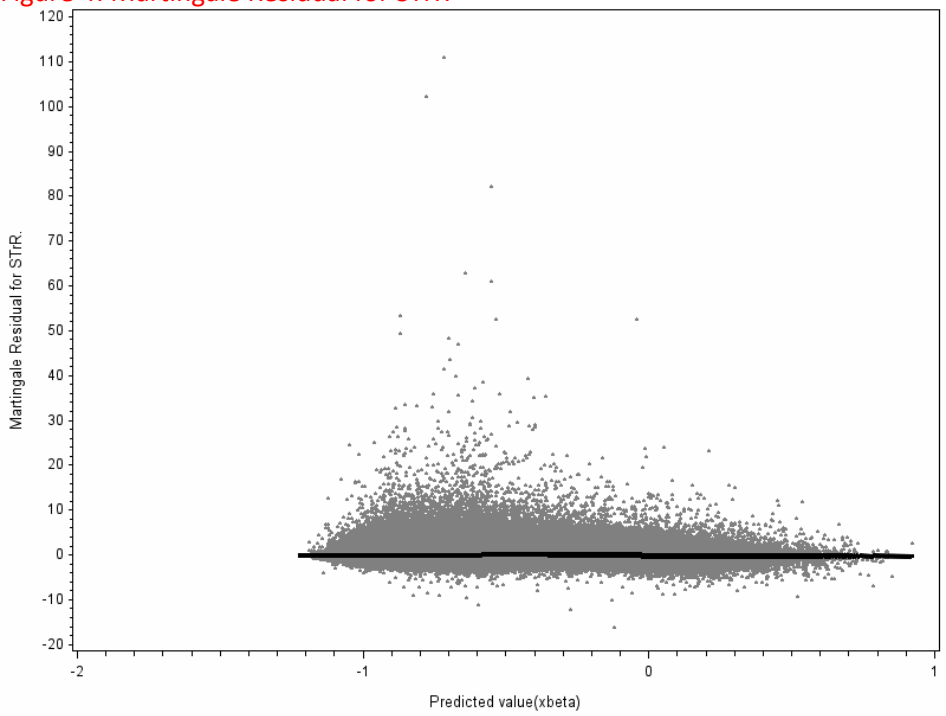
Reference: Barlow, W. E. and Prentice, R. L. (1988). Residuals for relative risk regression. Biometrika 75, 65{74.

Figure 3: Martingale Residual for STrR



For the Fall 2019 maintenance submission, the martingale residual plot is included below.

Figure 4: Martingale Residual for STrR



2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

The C-statistic for a recurrent event model measures the concordance between the observed rate of recurrent events and the model-based rate. The C-statistic for the STrR is 0.65.

For the Fall 2019 maintenance submission, the C-statistic for a recurrent event model measures the concordance between the observed rate of recurrent events and the model-based rate. The C-statistic for the STrR is 0.60.

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
We ranked each subject based on their average expected event rate. We then broke the subjects up into deciles and computed decile-specific observed and expected numbers of transfusions. Results are given in the table below; with the relative agreement between the observed and expected counts given in the last column. Overall, the model appears to have good calibration.

Table 23. Decile-specific observed and expected numbers of transfusions.

Decile	Observed transfusions	Expected transfusions	(Obs- Exp)/Exp
1	22042	22694.68	-0.029
2	24405	24611.55	-0.008
3	24232	24636.46	-0.016
4	24978	25427.46	-0.018
5	25507	26027.7	-0.020
6	26853	26851.19	0.000
7	27689	27377.81	0.011
8	28983	28324.41	0.023
9	31989	30352.24	0.054
10	40683	41057.5	-0.009

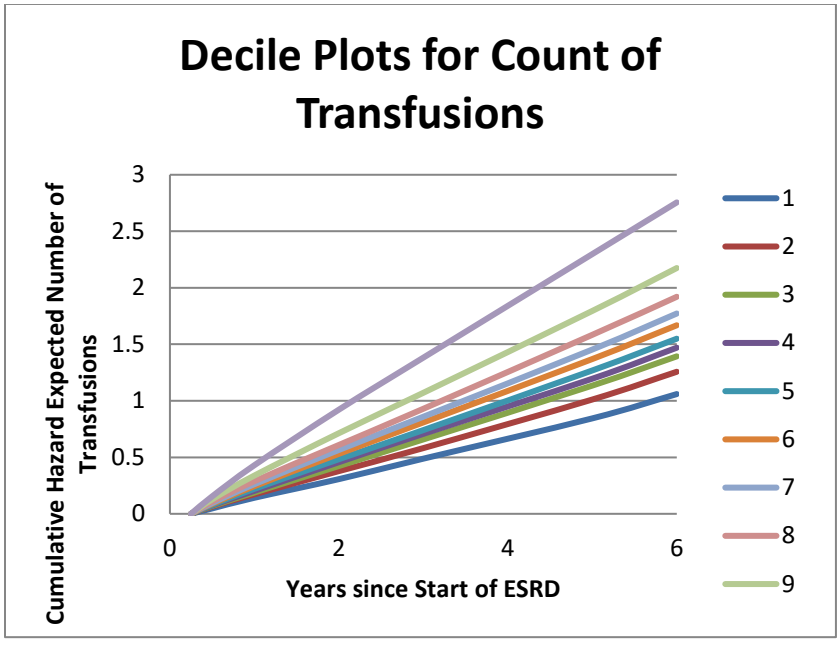
For the Fall 2019 maintenance submission,

Table 24. Decile-specific observed and expected numbers of transfusions.

Decile	Observed transfusions	Expected transfusions	(Obs- Exp)/Exp
1	25628	26291.28	-0.025
2	27058	27521.84	-0.017
3	27920	28248.66	-0.012
4	28287	28851.92	-0.020
5	29178	29427.17	-0.008
6	30209	30254.83	-0.002
7	31364	31327.19	0.001
8	32930	32630.66	0.009
9	37659	35771.26	0.053
10	56661	56569.19	0.002

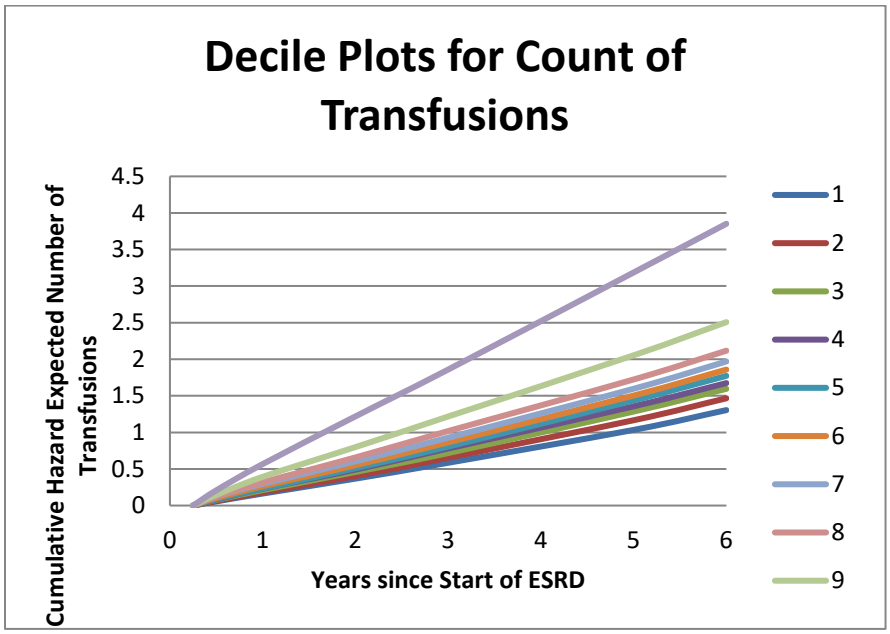
2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Decile plots (Figure 4) shows piecewise linear estimates of the cumulative rates by years since start of ESRD. The plot demonstrates that the risk factors in the model are discriminating well between patients. There 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).

Figure 5: Decile plots for count of transfusions.



For the Fall 2019 maintenance submission, we followed the methodology described in 2016.

Figure 6: Decile plots for count of transfusions.



2b3.9. Results of Risk Stratification Analysis:
N/A

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)
Covariates used as risk adjusters for STrR all have face and clinical validity and most of them are statistically significant at the 0.05 level. The residual plots show no lack of fit, while goodness-of-fit criteria show that there is added value in risk adjustment. The model appears to adequately discriminate the risk of transfusion among subjects; and, overall, is well-calibrated.

The interpretation from 2016 continues to apply to the Fall 2019 maintenance submission.

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

N/A

_____

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

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. 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), see 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 STrR.

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 describe 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 STrR 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.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)
The following table shows how the facilities are flagged for the year 2014, based on the method described above.

Table 25: Classification of Efron Empirical Null p-value for year 2014*.

Year 2014	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Better than expected	25	0.44	25	0.44%
As expected	5259	93.08	5284	93.52%
Worse than Expected	366	6.48	5650	100%

*Only for the facilities with patient years are greater than 10.

For the Fall 2019 maintenance submission, the following table shows how the facilities are flagged for the year 2017, based on the method described above under 2b4.1.

Table 26: Classification of Efron Empirical Null p-value for year 2017*.

Year 2017	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Better than expected	14	0.23%	14	0.23%
As expected	5,687	93.61%	5,701	93.84%
Worse than Expected	374	6.16%	6,075	100%

*Only for the facilities with greater than 10 patient years greater at risk.

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The results indicate that the STrR has the ability to classify facilities as being significantly better (or significantly worse) than expected; thereby demonstrating the ability to identify meaningful differences in the performance scores across facilities.

The interpretation for the Fall 2019 maintenance submission remains consistent with what was submitted in 2016.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic

missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The STrR measure is dependent on Medicare claims and other CMS administrative data for several important components of measure calculation, including identification of diagnoses that qualify for exclusion of the patient's time at risk from both the dialysis facility's observed transfusion events (numerator) as well as the facility's expected transfusion events calculation (denominator). In addition, inpatient and outpatient Medicare claims are the sole source for identification of transfusion events. For these reasons, STrR was originally developed and, subsequently implemented as, a measure limited to Medicare patients.

For STrR and several other Medicare-only measures developed by UM-KECC, the presence of active Medicare coverage has been defined using a combination of criteria including a defined minimum of paid claims for dialysis services and/or presence of a Medicare inpatient claim during an eligibility period. With the recent increase in Medicare Advantage (MA) coverage for Medicare chronic dialysis patients, and the systematic missing outpatient claims data for MA patients, these criteria have the potential to introduce significant bias into measure calculations that could affect results for dialysis facilities with either very low or high MA patient populations. As a result we decided to investigate the impact of MA missing data on STrR.

Medicare Advantage patient status was defined using Medicare Enrollment Database (EDB) criteria. Primary Medicare Fee for Service (FFS) coverage was identified using CMS administrative data, and active Medicare status utilized the combination of minimum dialysis paid claims and/or inpatient Medicare hospitalization claims briefly described above. Transfusion events were identified from Medicare claims from both inpatient and outpatient claims files using criteria defined in specifications (numerator details). For inpatient claims, separate analyses identified transfusion events from claims type "60" form Claims types "62-64", the claims type used for Medicare FFS and Medicare Advantage inpatient billing, respectively. For transfusion events for outpatient claims, we identified transfusion events from FFS and MA patients separately. Finally, we evaluated the source (inpatient vs. outpatient claims) for malignancy and other diagnosis categories separately for FFS and MA patients to determine 1) whether inpatient or outpatient claims files are the predominant source for excluding diagnoses, and 2) how the two coverage types influenced frequency of exclusion. Results of our investigation are summarized below.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Summary findings:

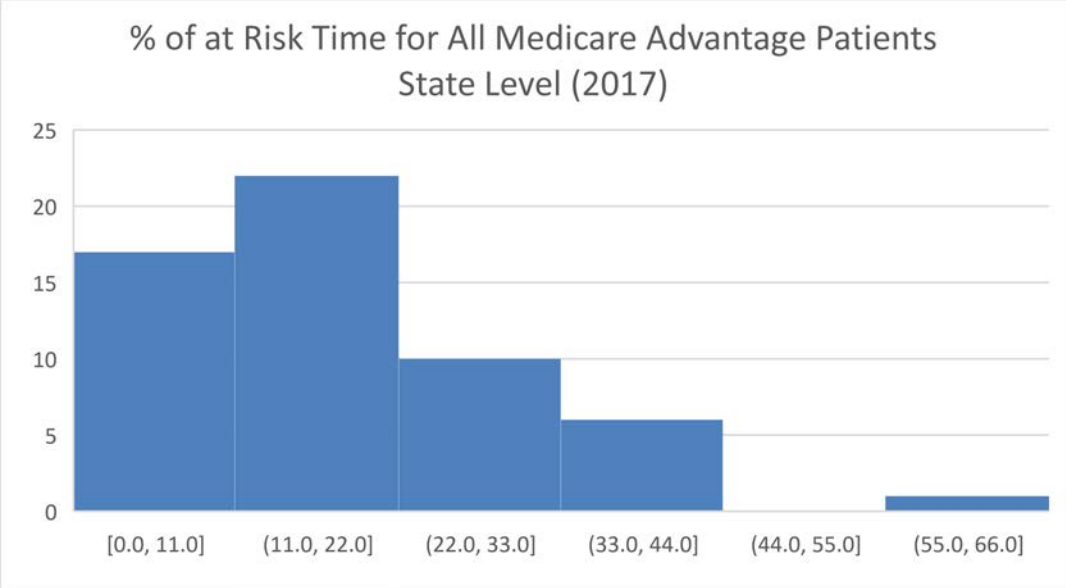
1. The percentage of patients with MA coverage receiving chronic dialysis in US dialysis facilities has approximately doubled in the last decade and is approaching 20% based on 2017 data.
2. When applied to MA patients, the historical definition of active Medicare coverage (described earlier) creates systematic bias in the STrR measure calculation through exclusion of MA patient time at risk in facilities unless the MA patient had one or more hospitalizations in the observation period. MA patients included because of hospitalization are very likely not representative of MA patients as a whole, instead reflecting a sicker subset. Calculating STrR using an alternative definition of time at risk for MA patients (using the Medicare EDB rather than inpatient or outpatient claims-based utilization), results in overall lower transfusion event rates compared to the original STrR approach. In addition, estimated risk of transfusion events for MA patients in risk-adjusted patient-level models is significantly lower than for other active Medicare patients when using the EDB to identify MA patient time at risk. This result is directionally opposite of the result for MA patients in models based on our original Medicare inpatient/outpatient claims-based inclusion criteria to determine time at risk for active Medicare patients. The model results in our investigative analyses are consistent with the expected lower transfusion rates for MA patients, based on the systematic missing outpatient claims for MA patients, a source for approximately 15% of transfusion events in the overall Medicare patient population.
3. STrR specifications call for exclusion of patients from facility-level time at risk for events if Medicare claims contain recent diagnosis of systemic malignancy or hereditary hemolytic anemia. Our analysis of 2014-17 data demonstrates that, for non-MA patients, 90% of possible excluding diagnoses were derived from outpatient claims types. For MA patients, only 22.3% of possible excluding diagnoses were from outpatient claims types. These results are consistent with the general unavailability of outpatient claims data for MA-associated services.

Table 27. Impact of different claim sources used to determine the exclusion criteria on time at risk for the STrR.

year	At risk time before exclusion	At risk time after inpatient claim only exclusion		At risk time after non-inpatient claim exclusion	
		count	% of exclusion	count	% of exclusion
2014-2017	540,726,906	502,953,822	7.0	449,515,907	16.9
2014	129,773,102	121,170,469	6.6	108,745,282	16.2
2015	133,256,649	124,255,888	6.8	110,208,214	17.3
2016	137,566,757	127,922,835	7.0	113,709,439	17.3
2017	140,130,398	129,604,630	7.5	116,852,972	16.6

Additional dialysis facility-level analyses (Figure 7) demonstrate a variable distribution of Medicare Advantage patient proportion following geographic boundaries. For example, the percentage of MA patient time at risk relative to total Medicare patient time at risk varies from a low of 2.2% in South Dakota to a high of 56.4% in Puerto Rico.

Figure 7: Distribution of the Percentage of Risk Time for All Medicare Advantage Patients at the State Level



Below, we compare the STrR results with and without the patient time at risk for Medicare advantage patients. The results are correlated at 0.94 ($p<.0001$).

Figure 8. Correlation between the STrR calculated with and without Medicare Advantage patients included.

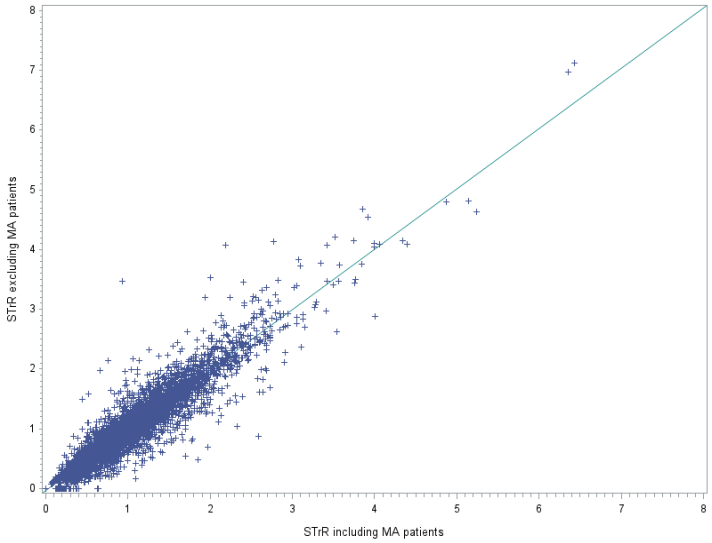


Table 28. Changes in flagging rates when excluding Medicare Advantage patients

STrR with MA patient years	STrR without MA patient years			
	Better than Expected	As Expected	Worse than Expected	Total
Better than Expected	9	18	0	27
	0.14	0.28	0	0.42
As Expected	4	5866	90	5961
	0.06	92.33	1.42	93.83
Worse than Expected	0	83	282	365
	0	1.31	4.44	5.75
Total	13	5967	372	6353
	0.2	93.92	5.86	100

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Based on the above results, we have excluded all Medicare Advantage at time at risk from the current version of STrR being submitted for maintenance endorsement. This minimizes risk of biased results at the dialysis facility level and is consistent with a number of other NQF-endorsed measures that are based on Medicare claims data.